



```

: | | | | |
IVTPI-----SVPAPV-----1079

AQHFQSVOSNAAYIQLHLCFGDNKIKABIRQGGIQLLVLLDHRMTEVHRSACGAL 616
: | | | | |
---QEPVLVDHSDVL-----TQDSELKRVSGVS-----HSAPSIV 1114

JYKANDDNKIALKNCGGIPALVRLLRKTTDTLEIRELVTVGLVNLNSCDALKMPIIQ 676
: | | | | |
---VESLTSMTPTIEL-----ACQTPASIGQ 1139

)ALAVLTNAVIIPIHSGWNSPQQDDRKICILHSSQVLRNATGCLRNVSAGAEARRMR 734
: | | | | |
)VIATAAHASLIENASVQSPSELDAETGL-----AGHEKLEALKMEQDR-- 1188

)GLTDALLYVIOGALGSHBIDSKTVENCVCILRNLSYRLAABTSOGOHMGTDLEDGL 794
: | | | | |
)DMGD-----DAIGTTTTDGG-----DBIPDITLTKGL- 1215

)ANGKDAESGCGWKKKKKQDQGVGPLDCAEBPKGIQMLWHPISVVKPYITLL 854
: | | | | |
)ALGKVIHA-----DGRETP--MPPD-----HPLDTDASTQQL 1248

)SNPDTLEGAGALONLAAGSKWKSVIIRAAVKEKGLPILVELLRINDNRVWCAV-A 913
: | | | | |
)SNPDVLITMSSAVEGASST-----MIE-----DIDASTSAVDA 1285

)RNWALDVRNKELLIGYAM-----RDLVRLPGGNNSNNTASK-AMSDDTVIACCTIL 967
: | | | | |
)NSMPFGAQNSTDQIPAAMTLSMDOECAQSMSTSTRNTTGTKLATFENLETASSTL 1345

)ITKNMENAKALRD-----AGIEKLVGISKSGDKHSPKVKVAAS- 1010
: | | | | |
II--RQPNAPSSRDETTAPMTPSFTNERIGGGGGGATFSIGTPPSHSPFVSECDY 1403
: | | | | |
KGQMDLESDPEVIQIVRHR-----MEQHKLLEKQORVEIKLR--SKIRVP 1450
: | | | | |
-SISPVTVSPN-----NRSASAPSPREMLSKERKTDVECTGSSNATYH 1100
: | | | | |
ISVNFPEMIGDDEADTTLALQSLAGNASJLSPASP-----PNTETTKVNTTIV 1501
: | | | | |
---KGEHTSRKDMATAQNTGISLTLYRNSYGAEDIKHNQVSAQVPQEPGRKQVETY 1157
: | | | | |
)VLATRMWSQSSTKSSNVSVSRHDNQSAPP---RHHHQHPH-PHPHPLQNH--Y 1555
: | | | | |
QONST 1164
PQNH 1562

```

09233857 US/

5555  
ATION:

Woman, Gregory

anagan, Peter  
NTION: HUMAN ORTHOLOGUES OF WART

E: 239/251

CATION NUMBER: US/09/233,857

G DATE: 1999-01-20

C DATE: 1008 01 21

DATE: 19  
ID NOS. 1

tSEO for Windows Version 3.0

;

VAN

Query Match	3.1%;	Score 194.5;	DB 4;	Length 1088;
Best Local Similarity	17.8%;	Pred. No. 2.5e-05;		
Matches 142;	Conservative 89;	Mismatches 242;	Indels 325;	
Qy	5	KPGGAFLGAMPVDP-----QPSSASEKTSLSLSPG-LNTSNGDGGSETE		
Db	66	KQMEPPPSASPADWIPPPPAIVGQFGAGSISVSGVGVGVGANGRVFKMM		
Qy	53	ASVKEQELQFERLITRELEAERQIVASQLERCKLGETSGSMSSMSAERQFOWQS		
Db	122	-----LMENKLIRKPSTERTASSHYLRCSALDSCAGSGSRSDS-----		
Qy	113	IEDELTTGLELYDSCIRSLQESGILDQDYSYNGRPSLLSQSALQLNKSPEGSEF		
Db	161	-----		
Qy	173	HSNQTLALGETTFSQLPARGTQARATGOSFGQTTSRAGHLAGPEPAP-----		
Db	165	HTHQP-----SSRTVGNPGNGGFSFSPSGSFSEV-----		
Qy	226	PREFPAPISGAFLHPDAPPAASAAAALYYSSSTLPAPP-----RGSGPLAA-----		
Db	200	PRNP7ASS--AATPPPPVPPTSCAVVKGRSPALNNRPPAIAPTQGRNSPVITQ		
Qy	272	-----POGGS-----PTKLORGSSA-----PEGATYAAPRGSS-----		
Db	258	QQQLTQQLKSLNLYPGGGSGAVVEPPPPVLYIQGGAGGAAPPPPPSYTASMQS-		
Qy	306	RLAKS---YSTSSPINIVVSSAGLSPIRTSPP-----		
Db	316	QSQSQDYRKSPSGSIYSATSGSPSFI TVSLPPFLAKPQPRVYQARSQQPFIIM		
Qy	341	ISSSPTHQLSSITGTVATLSPTKRLVHASEQYSKHSQELXATATLQRPGLSAAAG		
Db	376	QVQKPVLOTAVAPQSPSSASANSPVHVLAAPPSYQK--SAAVVQOQQQAAA-		
Qy	401	SOHGH-----LGPBLRALQSPHHIDP-----		
Db	431	QCHQHQSQKPATFTTTPPLVGLNLSKPNCLPSPSYAKSMOAKAATVVQOQQOQQQQ		
Qy	427	RVYQK-----		PPM
Db	491	QVQOQQOQQOQQQLQALRVLQAQARERDQRRERDQKQLANGNPGRQMLPPPPY		
Qy	441	QGDPLEPAAHTGT---YRTSTAPSPGVDSPVLQRTGSHQGPQNAAAAATFORASY		
Db	551	NSBIKEPPSCNNNNIQLSNSNLATTPTPIPPAKYNNNSNTGANSGGSGNSTGTT		
Qy	498	NYADPVRQIQCPSVESPSYKSGPALPPGTLIARSPSIDSTQKDPREF--GWRD		
Db	611	-----CKKI-----KHASFPIPERKKTISKKE-----EEERKEPRIRWAR		
Qy	556	IQMLQHQPFSVQSNAAAYLQHLCFGDNKTKABIRPQGGIQLLVLDLDRHMTVEVH		
Db	650	FKFMEQ--HIENVIKSYQR--TYRKNQLEKEMHKVG-----LPDQQTQIE--		
Qy	616	LRNLVYTKGANDDNKIALK 633		
Db	692	MRXNLNQK--ESNYIRLK 707		

Search completed: April 22, 2004, 12:15:56

Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

Gen search, using sw model

April 22, 2004, 12:06:19 ; Search time 64 Seconds  
(without alignments)  
5408.136 Million cell updates/sec

JS-09-501-171A-4

MFARXPGGAFLGAMPVDPQ.....PYSELNYETSHYPASPSWV 1225

3LOSUM62

Gapop 10.0 , Gapext 0.5

1586107 seqs, 282547505 residues

its satisfying chosen parameters: 1586107

angth: 0

angth: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

is the number of results predicted by chance to have a  
score than or equal to the score of the result being printed,  
divided by analysis of the total score distribution.

# SUMMARIES

%	Query	Match	Length	DB	ID	Description
100.0	1225	3	AAB07973			Aab07973 A human n
100.0	1233	4	ABG04996			Abg04996 Novel hum
96.1	1247	3	AAB07974			Aab07974 A murine
88.3	1084	2	AAY23900			Aay23900 Human res
85.1	1040	2	AAM24559			Aam24559 Presenili
61.4	756	2	AAM60664			Aam60664 Human ALA
44.2	1192	2	AAY23899			Aay23899 Human res
44.2	1211	2	AAM24560			Aam24560 Presenili
44.2	1211	3	AAY92336			Aay92336 Human p00
44.2	1211	5	ABE79001			Abb79001 Human p00
41.8	1110	4	AAE03648			Aae03648 Human ext
36.0	785	5	ABP41836			Abp41836 Human ova
35.9	696	4	ABG04995			Abg04995 Novel hum
26.7	1009	4	ABG20820			Abg20820 Novel hum
23.4	537	7	ADD27712			Add27712 Human adi
22.4	785	4	ABG20821			Abg20821 Novel hum
22.2	962	6	ADA27163			Ada27163 ARVCF pro
21.5	919	4	AAE10797			Aae10797 Human cat
21.4	941	7	ADB75256			Adb75256 Prostate
20.1	834	6	ADA54659			Ada54659 Human pro
18.4	779	4	ABB66318			Abb66318 Drosophal
13.1	837	4	AAU27825			Aau27825 Human ful
13.1	837	3	ABB36464			Aab36464 Human pla
12.4	472	6	ABU70901			Abu70901 Human adi
12.2	821	3	AAB36467			Aab36467 Xenopus l

## ALIGNMENTS

### RESULT 1

AAB07973

ID AAB07973 standard; protein; 1225 AA.

XX AC AAB07973;

XX DT 14-NOV-2000 (first entry)

XX DE A human neural plakophilin related armidillo protein.

XX KW Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; h

XX KW neural plakophilin related armidillo protein; Alzheimer's disea

XX KW Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke

XX KW multiple sclerosis; ischemia; stroke; neural paropathy; sciatic

XX KW motor neuron disease; peripheral neuropathy; neuropathy; diabet

XX KW spinal cord injury; facial nerve crush.

XX OS Homo sapiens.

XX PN WO200047615-A2.

XX PD 17-AUG-2000.

XX PF 11-FEB-2000; 2000WO-CA000126.

XX PR 12-FEB-1999; 99US-0119835P.

XX PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

XX PI St George- Hyslop PH, Fraser PE;

XX DR WPI; 2000-524531/47.

XX PS N-PSDB; AAA59700.

XX PT Stimulation of nerve cell growth using human Neural Plakophilin

XX PT Armidillo Protein (hNPRAP) polypeptide, useful for the treatmen

XX PT diseases such as Alzheimer's, Parkinson's, and stroke.

XX PS Claim 3; Page 24-27; 33pp; English.

XX CC The present sequence represents human Neural Plakophilin Relate

XX CC Armidillo Protein (hNPRAP) polypeptide. hNPRAP interacts with p

XX CC (PS) I and II (PS1 and PS2). The specification describes a meth

XX CC stimulating the growth of nerve cells, comprising contacting th

XX CC hNPRAP. The hNPRAP polypeptide and polynucleotide are useful fo

XX CC nerve damage caused by a variety of diseases or physical trauma

XX CC including Alzheimer's disease, Parkinson's disease, amyotrophic

XX CC sclerosis (ALS), multiple sclerosis, stroke, ischemia associate

26	770.5	12.1	797	3	AAB36462	Aab36462
27	770.5	12.1	797	4	AAM79226	Aam79226
28	770.5	12.1	797	6	ABU56576	Abu56576
29	766	12.0	797	3	AAB36466	Aab36466
30	719.5	11.3	726	3	AAB36463	Aab36463
31	719.5	11.3	726	6	ABU56444	Abu56444
32	702	11.0	817	4	AAM80210	Aam80210
33	684	10.7	425	3	AAB43829	Aab43829
34	581.5	9.1	444	4	ABG18885	Abg18885
35	475	7.4	86	4	AAM19695	Aam19695
36	475	7.4	86	4	ABB39418	Abb39418
37	475	7.4	86	4	AAB32941	Aab32941
38	475	7.4	86	4	ABB24194	Abb24194
39	475	7.4	86	4	AAM72710	Aam72710
40	475	7.4	86	4	AAM60099	Aam60099
41	475	7.4	86	4	ABG54408	Abg54408
42	475	7.4	86	5	ABG42534	Abg42534
43	452	7.1	198	6	ADA54263	Ada54263
44	407.5	6.4	1266	4	ABG10319	Abg10319
45	396	6.2	193	4	ABG20816	Abg20816

```

:5 AA:
100.0%; Score 6377; DB 3; Length 1225;
ilarity 100.0%; Pred.No.0;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ARKPGGAALGAMPVDPQPSASKTSLSPLQNTSNGDGETTETSAILASVKQOEL 60
|||||
ARKPGGAALGAMPVDPQPSASEKTSLSPLNTSNGDGETTETSAILASVKQOEL 60
|||||
ERLTRELEAERQIVASQLERCKLGETSGMSMSAEAEQFQWQSDGQKDIDELTG 120
|||||
ERLTRELEAERQIVASQLERCKLGETSGMSMSAEAEQFQWQSDGQKDIDELTG 120
|||||
LVDSCTRSLOESGILDPODYGTGERPSLLQSALQLNKSPEGSFOYPASVHNSQTLLAL 180
|||||
LVDSCTRSLOESGILDPODYGTGERPSLLQSALQLNKSPEGSFOYPASVHNSQTLLAL 180
|||||
TTTSQLPARGTQARATGQSFQGTTSRAGHLAGPEPPPPPPPPPPPPAPSLSAFL 240
|||||
TTTSQLPARGTQARATGQSFQGTTSRAGHLAGPEPPPPPPPPPPPPAPSLSAFL 240
|||||
APPAALAAALYYSSSTLPAPPRGSPLAAPCGSPKTLORGGSAPGATVAAAPRGSSP 300
|||||
APPAALAAALYYSSSTLPAPPRGSPLAAPCGSPKTLORGGSAPGATVAAAPRGSSP 300
|||||
SPSRLLAKSYSTSSPINIVVSSAGLSPIRVTSPTVQSTTSSSPIHQLSSTIGTVATLS 360
|||||
SPSRLLAKSYSTSSPINIVVSSAGLSPIRVTSPTVQSTTSSSPIHQLSSTIGTVATLS 360
|||||
XKRLVHASSEQYXKHSQELYATLQRPGLSIAAGRASVSSOHGHLGPELRALQSPHHI 420
|||||
XKRLVHASSEQYXKHSQELYATLQRPGLSIAAGRASVSSOHGHLGPELRALQSPHHI 420
|||||
IYEDRVYQKPMRSLQSQGDPLPPPAHTGYRTSTAPSSPGVDVSVPLQRTQSGHQPN 480
|||||
IYEDRVYQKPMRSLQSQGDPLPPPAHTGYRTSTAPSSPGVDVSVPLQRTQSGHQPN 480
|||||
IAATFORASYAAGPASNVADPYRQLQYCPVSESPYKSGPALPPEGTLARSPSIDSIQK 540
|||||
IAATFORASYAAGPASNVADPYRQLQYCPVSESPYKSGPALPPEGTLARSPSIDSIQK 540
|||||
PREFGWRDPELPEVIQMLQHOFPVSQSNAAAYLQHLCFQDNKIKAEIRQGGQLLVLDL 600
|||||
PREFGWRDPELPEVIQMLQHOFPVSQSNAAAYLQHLCFQDNKIKAEIRQGGQLLVLDL 600
|||||
DHRMTEVHRSACGALRNLYVGKANDDNKIALKNCGGIPALVRLIRKTTDLEIRLVTVG 660
|||||
DHRMTEVHRSACGALRNLYVGKANDDNKIALKNCGGIPALVRLIRKTTDLEIRLVTVG 660
|||||
WNLSLSDCALKNPIIQDALAVLTNAVIIIPHSGWENSPLODRKTLQIHSSQVLRNATGCLR 720
|||||
WNLSLSDCALKNPIIQDALAVLTNAVIIIPHSGWENSPLODRKTLQIHSSQVLRNATGCLR 720
|||||
IVSSAGEEAPRRMRECDGLTDALLYVTQSALGSSEIDSKTVENCVCILRNLSYRLAAETS 780
|||||
IVSSAGEEAPRRMRECDGLTDALLYVTQSALGSSEIDSKTVENCVCILRNLSYRLAAETS 780
|||||
XQGHMGTDELGLLGCGBANGKDAESSGCWKKKKKKSSQDQWDGVGPLPDCAEPPKGIOM 840
|||||
XQGHMGTDELGLLGCGBANGKDAESSGCWKKKKKKSSQDQWDGVGPLPDCAEPPKGIOM 840
|||||
JMHPSIVKPYLTLLSECSNPDTLEAGALQNLAAAGSNKWSYVIRAAVTRKEKGLPTLVEL 900
|||||
JMHPSIVKPYLTLLSECSNPDTLEAGALQNLAAAGSNKWSYVIRAAVTRKEKGLPTLVEL 900
|||||
LRINDNRVCAVATALRNALDVRNKELIGKYAMPDLVHRLPGGNNSNNTASKMSDDTV 960
|||||
LRINDNRVCAVATALRNALDVRNKELIGKYAMPDLVHRLPGGNNSNNTASKMSDDTV 960
|||||

```

forensics, gene mapping, identification of mutations or genetic disorders or other traits to assess biodiversity or other types of data and products dependent on DNA and sequences. AB00010-AB030377 represent novel human diagnostic sequences of the invention. Note: the sequence data for this t appear in the printed specification, but was obtained in rmat directly from WIPO at pub/published\_pct\_sequences

AA;  
 100.0%; Score 6377; DB 4; Length 1233;  
 arity 100.0%; Pred. No. 0;  
 conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 KPPGAAPLGAMPVDPQSSASEKSSLSPLGNTSNGDSETTTSAIIASVKEQL 60  
 KPPGAAPLGAMPVDPQSSASEKSSLSPLGNTSNGDSETTTSAIIASVKEQL 68  
 LTRLEAERQIVASQLERCKLGSETGSMSSSAEEQFQWQSQDQKDIEDELTTG 120  
 LTRLEAERQIVASQLERCKLGSETGSMSSSAEEQFQWQSQDQKDIEDELTTG 128  
 DSCIRSQESGILDPQDYSTGERPSLSQSALQLNSKPEGSFOYPASVHSNQTAL 180  
 DSCIRSQESGILDPQDYSTGERPSLSQSALQLNSKPEGSFOYPASVHSNQTAL 188  
 PSQLPARGTQARATQOSQCTTSRACHLAGPEPAPPPPPPPPPPPPPPPPPPP 240  
 PSQLPARGTQARATQOSQCTTSRACHLAGPEPAPPPPPPPPPPPPPPPPPPP 248  
 PAAAAALYSSSTLPAPRGGSPLAAPQGGSPKLPQGGSAPEGATVAAPRGSSP 300  
 PAAAAALYSSSTLPAPRGGSPLAAPQGGSPKLPQGGSAPEGATVAAPRGSSP 308  
 SRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHOLSSTIGYATLS 360  
 SRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHOLSSTIGYATLS 368  
 LVHASEQYKHSQELIYATLQPGSLAAGSRASVSSQHGHLGPELRALQPEHHI 420  
 LVHASEQYKHSQELIYATLQPGSLAAGSRASVSSQHGHLGPELRALQPEHHI 428  
 EDVYQKPMRSLQSQDPLPPTAHTGYRTSTAPSSPGVDSVPLQRTSQHGPN 480  
 EDVYQKPMRSLQSQDPLPPTAHTGYRTSTAPSSPGVDSVPLQRTSQHGPN 488  
 TFQASAAAGPASNYADPYRQIQCPSVESPYSGSGPALPPEGTLARSPDSIQK 540  
 TFQASAAAGPASNYADPYRQIQCPSVESPYSGSGPALPPEGTLARSPDSIQK 548  
 FGWRDPPELPEVIMLOHOPPSVQSNAAAYLOHLCFGDNKIKAEIRROGGIQLVDL 600  
 FGWRDPPELPEVIMLOHOPPSVQSNAAAYLOHLCFGDNKIKAEIRROGGIQLVDL 608  
 MTEVHRSACGALNLYVGKANDDNKIALKNCGGIPALVRLLRKTTDLRELVTVG 660  
 MTEVHRSACGALNLYVGKANDDNKIALKNCGGIPALVRLLRKTTDLRELVTVG 668  
 SSCDALKMPITQDALAVTNNAVITPHSGWENSPLQDDRKIQLHSSQVLRNATGCLR 720  
 SSCDALKMPITQDALAVTNNAVITPHSGWENSPLQDDRKIQLHSSQVLRNATGCLR 728  
 IAGEEARRRMRCDGTLTDLVYIQSALGSSEIDSKTNCVCLNLSVRLAAETS 780  
 IAGEEARRRMRCDGTLTDLVYIQSALGSSEIDSKTNCVCLNLSVRLAAETS 788  
 IMGTDELGLLCCGANGKDAESGCGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 840  
 IMGTDELGLLCCGANGKDAESGCGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 848  
 STIVKPYLTLLESCSNPDTLEGAAGALQNLAAAGSWKSVYIRAAVVRKXGLPILVEL 900

Db 849 LWFPSIVKPYLTLLESCSNPDTLEGAAGALQNLAAAGSWKSVYIRAAVVRKXGLP;  
 Qy 901 LRINDRVVCAVATARNMALDVNRKELIGYAMRDLVHRLPGGNNNSNTASKAM;  
 Db 909 LRINDRVVCAVATARNMALDVNRKELIGYAMRDLVHRLPGGNNNSNTASKAM;  
 Qy 961 TAVCCTLHEVITKMNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLNS;  
 Db 969 TAVCCTLHEVITKMNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLNS;  
 Qy 1021 DLRSLYKDGWSQYHFVASSSTIERDRQRPYSSSRTSPSISPVRSPPNRSASAPAK;  
 Db 1029 DLRSLYKDGWSQYHFVASSSTIERDRQRPYSSSRTSPSISPVRSPPNRSASAPAK;  
 Qy 1081 ISLKERKTDYECTGSNATYHGAKGHTSRKDAMTAQNTGISTLYRNSYGAPEADII;  
 Db 1089 ISLKERKTDYECTGSNATYHGAKGHTSRKDAMTAQNTGISTLYRNSYGAPEADII;  
 Qy 1141 SAQVPQPSRKDYETQYQFQNSTRNVDSPFQVHHRPPASEYTMHLGLKSTGI;  
 Db 1149 SAQVPQPSRKDYETQYQFQNSTRNVDSPFQVHHRPPASEYTMHLGLKSTGI;  
 Qy 1201 YSAARPYSELNYETSHYPASPSWSV 1225  
 Db 1209 YSAARPYSELNYETSHYPASPSWSV 1233  
 RESULT 3  
 AAB07974  
 ID AAB07974 standard; protein; 1247 AA.  
 XX AC AAB07974;  
 XX DT 14-NOV-2000 (first entry)  
 DE A murine neural plakophilin related armidillo protein.  
 KW Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hN1  
 KW neural plakophilin related armidillo protein; Alzheimer's disease;  
 KW Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;  
 KW multiple sclerosis; ischemia; stroke; neural paropathy; sciatic;  
 KW motor neuron disease; peripheral neuropathy; neuropathy; diabetic;  
 KW spinal cord injury; facial nerve crush.  
 XX Mus sp.  
 OS WO2000047615-A2.  
 XX PN 17-AUG-2000.  
 XX PD 11-FEB-2000; 2000WO-CA000126.  
 XX PF 12-FEB-1999; 99US-0119835P.  
 XX PR (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 XX PA St George- Hyslop PH, Fraser PE;  
 XX PI WPI: 2000-524531/47.  
 XX DR N-PSDB; AAA59701.  
 XX PT Stimulation of nerve cell growth using human Neural Plakophilin I  
 PT Armidillo protein (hNPRAP) polypeptide, useful for the treatment  
 PT diseases such as Alzheimer's, Parkinson's, and stroke.  
 XX PS Disclosure; Page 30-33; 33pp; English.  
 XX CC The present sequence represents Neural Plakophilin Related Armid:  
 CC Protein (NPRAP) polypeptide. Human NPRAP interacts with presenil:  
 CC I and II (PS1 and PS2). The specification describes a method for  
 CC stimulating the growth of nerve cells, comprising contacting ther  
 CC hNPRAP. The hNPRAP polypeptide and polynucleotide are useful for  
 CC nerve damage caused by a variety of diseases or physical traumas.

zheimer's disease, Parkinson's disease, amyotrophic lateral  
 (S), multiple sclerosis, stroke, ischemia associated with  
 al parapathy, motor neuron diseases, sciatic crush,  
 neuropathy, neuropathy associated with diabetes, spinal cord  
 facial nerve crush

7 AA;

96.1%; Score 6129; DB 3; Length 1247;

larity 94.8%; Pred. No. 0;

Conservative 8; Mismatches 29; Indels 28; Gaps 3;

RKPPGAALGAMPVDPQSSASEKTSISPLNTSNGDGETETTSAILASVKEQEL 60

RKQSGAALGAMPVDPQSSASEKNSLSPLNTSNGDGETETTSAILASVKEQEL 60

RLTRELEAERQIVASQLERCKLGSETGSMSSSABEFOQSQDQKQDELTG 120

RLTRELEAERQIVASQLERCKLGSETGSMSSSABEFOQSQDQKQDELTG 120

VDSICIRLSQESGILDPQDYSTGERPSLLSQALQNSKPEGSFOYPASVHNSQTLAL 180

VDSICIRLSQESGILDPQDYSTGERPSLLSQALQNSKPEGSFOYPASVHNSQTLAL 180

TPSQLPARQTQARATQCSFQSQTTSRAGLAGPEAPPPPPPPPPPPPPPPPSLGSAPHL 240

APSQLPARSTQARAAGQSFQSQTTGRAGLAGSEPA-PPPPPPPPPPPPPPPSLGSAPHL 239

PPAAAAALYSSSTLPAPPRGSGPLAAPQGGSPKLOGGSAPEGATYAAPRGSSP 300

APP--AAALYSSSTLPAPPRGSGPLTTTQGGSPKLOGGSAPEGATYAAPRGSSP 297

PSRLAKSYSTSPINIVSSAGLSPIRVTSPPTVQSTISSPIHOLSTIGTYATLS 360

PSRLAKSYSTSPINIVSSAGLSPIRVTSPPTVQSTISSPIHOLSTIGTYATLS 357

RLVHASEQYKHSQELYTATLQRPGLSAAAGSRASYSSQHGHLGPELALQSPEHHI 420

RLVHASEQYKHSQELYTATLQRPGLSAAAGSRASYSSQHGHLGPELALQSPEHHI 417

CYEDRVYQPPMPSLSQSDPLPPAHTGYRTSTAPSGVDSPVLTQSGHQGPON 480

CYEDRVYQPPMPSLSQSDPLPPAHTGYRTSTAPSGVDSPVLTQSGHQGPON 477

YATFORASYAAGPASYADPYRQYQYCPVSEPSYSGSPALPPEGLTARSPTSISQK 540

YATFORASYAAGPASYADPYRQYQYCPVSEPSYSGSPALPPEGLTARSPTSISQK 537

REFGWRDPPELPEVIQMLHQFPVQSNAAAYLQHLGFGDNKIKAEIRROGGIQLVYL 600

REFGWRDPPELPEVIQMLHQFPVQSNAAAYLQHLGFGDNKIKAEIRROGGIQLVYL 597

IRMTVEHRSACGALRNLYVGKANDDNKIALKNCGGIPALVRLRKTDLLEIRELVTGV 660

IRMTVEHRSACGALRNLYVGKANDDNKIALKNCGGIPALVRLRKTDLLEIRELVTGV 657

JLSSCDALXMPITQDALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLR 720

JLSSCDALXMPITQDALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLR 717

SSAGEARRMRRECDGLTDALLVVIOSALGSSSIDSKTVENCVCILRNLSYRLAAETS 780

SSAGEARRMRRECDGLTDALLVVIOSALGSSSIDSKTVENCVCILRNLSYRLAAETS 777

JHMGTDRLDGLGCEANGDAESSGGWKKKKKKKQDQWDGVGVLPCDAEPPKGIQM 840

JHMGTDRLDGLGCEANGDAESSGGWKKKKKKKQDQWDGVGVLPCDAEPPKGIQM 837

FPISTVKKPYLTLSECSNPDTLEGAAGALQNLAAAGSWK----- 879

FPISTVKKPYLTLSECSNPDTLEGAAGALQNLAAAGSWK----- 897

--WSVYIRAAVRKEKGLPILVELLRIDNDRVVCVATALRNALDVRNKKELICKYMR 935

898 CLPQWSVYIRAAVRKEKGLPILVELLRIDNDRVVCVATALRNALDVRNKKELI  
 936 DLVHRLPGGNNNSNTASKAMSDDTVTAVCCITLHEVITKMNENAKALRDAGGIEKI  
 958 DLVHRLPGGNNNSNGSKAMSDDTVTAVCCITLHEVITKMNENAKALRDAGGIEKI  
 996 SKGDKHSPKVVVKAASQVLNSMWQYRDLRSLYKKDGWSQYHFVASSSTIERDRQRI  
 1018 SKGDKHSPKVVVKAASQVLNSMWQYRDLRSLYKKDGWSQYHFVASSSTIERDRQRI  
 1056 TPSISPVVRVPPNRRSASAPASPREMISLKDKRKYDCTGSGNATYHGAKGHTSRI  
 1078 TPSISPVVRVPPNRRSASAPASPREMISLKDKRKYDCTGSGNATYHGAKGHTSRI  
 1116 QNTGISTLYRNSVGCAPAEEDIKHQVSAQPVQPEPSRKDYETQYQFQNSYTRYDEK  
 1138 QNTGISTLYRNSVGCAPAEEDIKHQVSAQPVQPEPSRKDYETQYQFQNSYTRYDEK  
 1176 VHRPPASEYTMHLGLKSTGNYVDYFSAARPYSSELNYTSHYPASPDQSV 122  
 1198 VHRPPASEYTMHLGLKSTGNYVDYFSAARPYSSELNYTSHYPASPDQSV 124

## RESULT 4

AAV23900

ID AAV23900 standard; protein; 1084 AA.

XX AC AAV23900;

XX DT 27-SEP-1999 (first entry)

XX DE Human reseinlin binding armadillo protein GT24/hnPRAP.

XX KW Human; presenilin 1; P51; presenilin-binding protein; interacti

XX KW presenilin allele; Alzheimer's disease; senile dementia;

XX KW psychiatric disease; schizophrenia; depression; neurological di

XX KW stroke; cerebral haemorrhage; p0071; armadillo protein.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 115 /note= "not specified"

XX FT WO9935501-A1.

XX PD 15-JUL-1999.

XX PF 08-JAN-1999; 99WO-CA000018.

XX PR 09-JAN-1998; 98US-0070948P.

XX PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

XX PI St George- Hyslop PH, Fraser PE;

XX DR WPI; 1999-419410/35.

XX PT Identifying substances that alter presenilin interactions, usef

XX PT screening individuals for presenilin alleles associated with Al

XX PT disease - useful for diagnosis of Alzheimer's disease.

XX PS Claim 9; Fig 1; 40pp; English.

XX CC The present sequence represents a human presenilin-binding protei

XX CC termed GT24/human neuronal plakoglobin related armadillo protein

XX CC hnPRAP). The specification describes a method for identifying su

XX CC that alter the interaction of a presenilin with a presenilin-bi

XX CC protein. The method comprises contacting the interacting domain

XX CC presenilin protein to a presenilin-binding protein in the prese

XX CC test substance, and measuring the interaction of the presenilin

XX CC presenilin-binding protein. The method can be used to screen in

n alleles associated with Alzheimer's disease and related  
ch as senile dementia's, psychiatric diseases such as  
and depression, and neurological disease, such as stroke  
haemorrhage

AA;

88.3%; Score 5633; DB 2; Length 1084;

arity 99.8%; Pred. No. 0;

conservative 0; Mismatches 2; Indels 0; Gaps 0;

SQSALQNSKPEGSFQYPASVHSNQTALGETTSPQLPARCTQARATGQSFSQGT 207

SQSALQNSKPEGSFQYPASVHSNQTALGETTSPQLPARCTQARATGQSFSQGT 66

HLAGEPAPPP 267

HLAGEPAPPP 126

PQGSPTKLQSGSAPEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVSSAGLS 327

PQGSPTKLQSGSAPEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVSSAGLS 186

TSPTVQSTISSPPIHQLSSTIGYATLSPTKRLVHASEQYKSHQELYATATLQR 387

TSPTVQSTISSPPIHQLSSTIGYATLSPTKRLVHASEQYKSHQELYATATLQR 246

AAGRASVSSOHGHLGPELALQSPHEHIDPIYEDRVYQKPPMSLSQSGDPLPP 447

AAGRASVSSOHGHLGPELALQSPHEHIDPIYEDRVYQKPPMSLSQSGDPLPP 306

TYRTSTAPSSPGVDSVPLQRTSGHQGPQNAATAFQRAASYAAGPASNYADPYRLQ 507

TYRTSTAPSSPGVDSVPLQRTSGHQGPQNAATAFQRAASYAAGPASNYADPYRLQ 366

VESPYKSGPALPEGTTLARSPIDSIOKDPREFGWRDPPELPIVQLMHOFPSPVQ 567

VESPYKSGPALPEGTTLARSPIDSIOKDPREFGWRDPPELPIVQLMHOFPSPVQ 426

AYLOHLCFGDNKIKAEIRROGGIQLLVLLDHRMTEVHRSACGALRNLYVGKANDD 627

AYLOHLCFGDNKIKAEIRROGGIQLLVLLDHRMTEVHRSACGALRNLYVGKANDD 486

LKNCGGIPALVRLKRTDLEIRELVGTGLVNLSSCDALXMPHIIQDALAVLTNAV 687

LKNCGGIPALVRLKRTDLEIRELVGTGLVNLSSCDALXMPHIIQDALAVLTNAV 546

IGWENSPLODDRKIQLHSSQVLNATGCLRNVSAGEEARRMRRECDGLTDALLY 747

IGWENSPLODDRKIQLHSSQVLNATGCLRNVSAGEEARRMRRECDGLTDALLY 606

GSSEIDSKTVENCVCILRNLSVRLAAETSGQGHMGTDELGLLCEANGKDAESSG 807

GSSEIDSKTVENCVCILRNLSVRLAAETSGQGHMGTDELGLLCEANGKDAESSG 666

TKKKKKKSDQMDGVGFLPDCABPPKGIQMLMHPISIVKPYLTLLSCSNPDTLEGAA 867

TKKKKKKSDQMDGVGFLPDCABPPKGIQMLMHPISIVKPYLTLLSCSNPDTLEGAA 726

NLAAGSWKSVYVIRAARVKEKGLPILVELLRIDNDRVVCATARNALRNALDVRNKE 927

NLAAGSWKSVYVIRAARVKEKGLPILVELLRIDNDRVVCATARNALRNALDVRNKE 786

CYAMRDLVHRLPGNNNNNTASKAMSDDTVTAVCCITLHEVITKMNENAKALRDAGGI 987

CYAMRDLVHRLPGNNNNNTASKAMSDDTVTAVCCITLHEVITKMNENAKALRDAGGI 846

IGISKSGDKHSPKVKAAASQVLNSMWQYRDLRSKYKCGWSQYHFVASSSTIERDR 1047

IGISKSGDKHSPKVKAAASQVLNSMWQYRDLRSKYKCGWSQYHFVASSSTIERDR 906

YSSRTPTSPVVRVSPNNRSASAPSPREMI SLKERTDYECTGSGNATHYGHGFI 1107

907 QRPVSSRTPTSPVVRVSPNNRSASAPSPREMI SLKERTDYECTGSGNATHYGHGFI

1108 SRKDAMTAQNTIGISTLYRNSYGAPAEIDIKHNQVSAQPVPQEPSPKDKYETQYQFQNI

967 SRKDAMTAQNTIGISTLYRNSYGAPAEIDIKHNQVSAQPVPQEPSPKDKYETQYQFQNI

1168 DESFFEDQVHRRPPASBYTMHLGLKSTGNYVDYFSAARPPYSELNAYETSHYPASPD

1027 DESFFEDQVHRRPPASBYTMHLGLKSTGNYVDYFSAARPPYSELNAYETSHYPASPD

RESULT 5

AAW24559

ID AAW24559 standard; protein; 1040 AA.

XX AAW24559;

AC AAW24559;

DT 06-FEB-1998 (first entry)

XX Presenilin-interacting protein GT24.

DE Presenilin-interacting protein; human; Alzheimer's disease; diagr

KW therapy; transgenic animal; animal model; GT24.

XX Homo sapiens.

XX Key

FH Misc-difference 12

FT /note= "encoded by GST"

FT Misc-difference 40

FT /note= "encoded by CCR"

FT Misc-difference 71

FT /note= "encoded by AKC"

FT 346. .862

FT /note= "presenilin-interacting domain"

XX WO9727296-A1.

XX 31-JUL-1997.

XX 27-JAN-1997; 97WO-CA000051.

XX 26-JAN-1996; 96US-00592541.

XX 05-JUL-1996; 96US-0021673P.

XX 12-JUL-1996; 96US-0021700P.

XX 08-NOV-1996; 96US-0029895P.

XX 02-JAN-1997; 97US-0034590P.

XX (HSCR-) HSC RES &amp; DEV LP.

XX (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

XX St George- Hyslop PH, Frazer PE, Rommens JM;

XX WPI: 1997-393684/36.

XX N-PSDB; AAT79967.

XX Presenilin-interacting protein genes - used to develop products

XX diagnosis, therapy and study of Alzheimer's disease and related

XX disorders.

XX Claim 1; Page 99-101; 133pp; English.

XX GT24, a human protein with 'armadillo' repeats, has been identifi

XX to screen a human brain cDNA library for clones which interacted

XX presenilin transmembrane 6-7 loop domain; mutations in this loop

XX are known to be causative of Alzheimer's disease (AD). 9 PIP gen

XX sequences (see AAT79966-74) including G124 were identified. PIP:

XX acids, PIP proteins and peptides (especially the presenilin inte

XX domain), antibodies to PIPs, cells transformed with PIP nucleic

XX and transgenic animals altered with PIP nucleic acids can be use

XX diagnosis, therapy and study of AD and related disorders. They c

XX used to identify compounds which can modulate the expression of

ch bind to a PIP or modulate its activity

10 AA;

85.1%; Score 5426; DB 2; Length 1040;

Larity 99.6%; Pred. No. 0;

Conservative 0; Mismatches 4; Indels 0; Gaps 0;

.PARGTQARATGSGFSGTTSRAGHLAGBPAPPPPPPPPPPPPPPSLGSFAFHLDPDAPP 245

.PARGTQARATGSGFSGTTSRAGHLAGBPAPPPPPPPPPPPPPPSLGSFAFHLDPDAPP 60

AAALYSSSTLPAPRGGSPPLAAGPGSTKLORGSAPEGATYAAPRGSSPKQSPS 305

AAALYSSSTLPAPRGGSPPLAAGPGSTKLORGSAPEGATYAAPRGSSPKQSPS 120

AKSYSTSSPINIVSSAGLSPIRVPTVQSTISSPIHQLSSTIGTATLSPTKRL 365

AKSYSTSSPINIVSSAGLSPIRVPTVQSTISSPIHQLSSTIGTATLSPTKRL 180

ISEQYKHSQELVATATLQRPGLAGSRASYSSQHGHLGPELRALQSPHHIDPIYE 425

ISEQYKHSQELVATATLQRPGLAGSRASYSSQHGHLGPELRALQSPHHIDPIYE 240

LYOKPMRSLUSQSGPPLPAHTGTVRTSTAPSSPGVDSVPLQRTGSHQGPQNAAT 485

LYOKPMRSLUSQSGPPLPAHTGTVRTSTAPSSPGVDSVPLQRTGSHQGPQNAAT 300

YASYAAGPASNYADPYRQLQYCFVSFYSKSGPALPPEGTLARSPIDSQKDPREF 360

YASYAAGPASNYADPYRQLQYCFVSFYSKSGPALPPEGTLARSPIDSQKDPREF 545

YDEPDEVIMLOHQPPSVOSNAAYLQHLCFGDNKIKABIRQGGIQLLVLDLDRM 605

YDEPDEVIMLOHQPPSVOSNAAYLQHLCFGDNKIKABIRQGGIQLLVLDLDRM 420

YHSACGALRNLYVYKANDDNKIALKNCGGIPALVRLRKTDTLEIRLVTVGLWNLS 665

YHSACGALRNLYVYKANDDNKIALKNCGGIPALVRLRKTDTLEIRLVTVGLWNLS 480

YALKMPIIQDALAVLTNAVYIIPHSWENSPLODDRKIQLHSSQVLRNATGCLRNVSSA 725

YALKMPIIQDALAVLTNAVYIIPHSWENSPLODDRKIQLHSSQVLRNATGCLRNVSSA 540

YARRRRECGLTDALLYVYQSALGSEIDSKTVENCVCILRNLSYRLAETSGQGHM 785

YARRRRECGLTDALLYVYQSALGSEIDSKTVENCVCILRNLSYRLAETSGQGHM 600

YELDGLLCCGANGKDAESSGCGWKKKKKSKQDQWGVPLPDCAPPKGIQMLWHP 845

YELDGLLCCGANGKDAESSGCGWKKKKKSKQDQWGVPLPDCAPPKGIQMLWHP 660

YLYTLTLLSECSNPDTLEGAGALQNLAAAGSKWSVYIRAAVRKEKGLPILVELLRIDN 905

YLYTLTLLSECSNPDTLEGAGALQNLAAAGSKWSVYIRAAVRKEKGLPILVELLRIDN 720

YVCATATARNMALDVNRKELIGYAMRDVLVHRLPGGNNSNTASKAMSDDTVTAVCC 965

YVCATATARNMALDVNRKELIGYAMRDVLVHRLPGGNNSNTASKAMSDDTVTAVCC 780

YEVITKNENAKALRDAGGIEKLVGISKSGDKSPKVVKAASQVLNSMWMQYDLRSL 1025

YEVITKNENAKALRDAGGIEKLVGISKSGDKSPKVVKAASQVLNSMWMQYDLRSL 840

YKQWSQYHVFVASSSTIERDRQRPVSSRTSPISFVRVSPNRRSASAPASPREMISLKE 1085

YKQWSQYHVFVASSSTIERDRQRPVSSRTSPISFVRVSPNRRSASAPASPREMISLKE 900

YDVECTGSNATYHGKGEHTSRKDMATQNTGISTLYRNSYGAPAEIKKNQVSAQPV 1145

YDVECTGSNATYHGKGEHTSRKDMATQNTGISTLYRNSYGAPAEIKKNQVSAQPV 960

YSPRKKDYETQPFQNFQSTRNYDESFEDQVHRRPPASEYTMHLGLKSTGNTVDFYSAAR 1205

Db 961 PQEPRKDYETQPFQNFQSTRNYDESFEDQVHRRPPASEYTMHLGLKSTGNTV

QY 1206 PYSELNYETSHYPASPDWSV 1225

Db 1021 PYSELNYETSHYPASPDWSV 1040

# RESULT 6

AAW60664

ID AAW60664 standard; protein; 756 AA.

XX AAW60664;

DT 26-OCT-1998 (first entry)

XX Human ALARM protein.

XX ALARM; adherens-junction linked arm protein; delta-catenin; pre diagnosis; Alzheimer's disease; antibody; human.

XX Homo sapiens.

Key Location/Qualifiers

FT Region 159..202

FT Region /note= "arm repeat i"

FT Region 203..243

FT Region /note= "arm repeat ii"

FT Region 412..450

FT Region /note= "arm repeat iii"

FT Region 459..499

FT Region /note= "arm repeat iv"

XX WO9825142-A1.

XX 11-JUN-1998.

XX 02-DEC-1997; 97WO-US022093.

XX 02-DEC-1996; 96US-0031556P.

XX (BGM ) BRIGHAM & WOMENS HOSPITAL.

XX Kosik KS, Zhou J;

XX WPI; 1998-333494/29.

XX N-PSDB; AAV38305.

XX Human ALARM polypeptide(s) binding to presenilin 1 - useful, e. detect presenilin 1 and diagnose human diseases caused by mutant protein interacting with ALARM.

XX Claim 4; Fig 1A-F; 62pp; English.

XX This novel human protein, named ALARM or delta-catenin, is able to presenilin 1. Its amino acid sequence was deduced from a cDN. (see AAV38305) isolated from a human brain cDNA library. ALARM adherens-junction linked arm protein) is expressed almost exclu brain tissue. It contains 4 copies of the arm repeat that is characteristic of catenin proteins. Overall, ALARM and delta-ca 52.3% similar and 32.1% identical. The invention provides iso polymers, nucleic acids, vectors, host cells and specific an Presenilin 1 protein can be detected in samples (especially hum cerebrospinal fluid), by contacting with ALARM (claimed). Mutat: presenilin 1 gene is associated with familial Alzheimer's disea can be used to diagnose human diseases caused by a mutant form protein interacting with ALARM, by analysing fluid samples to d ALARM-interacting protein (claimed). ALARM can also be used to altered levels of presenilin 1, by contacting a sample with ALA determining bonding (claimed). ALARM polypeptides and antibodie used to detect an ALARM-containing complex in biological sample: contacting with polypeptide or antibody and determining whether polypeptide/antibody binds to a sample component (claimed). The



e also useful to detect ALARM polypeptides and to measure candidate compounds on expression or localisation of ALARM. ul to evaluate engineered cells prior to introduction in to inhibit abnormal ALARM activity or to generate antibodies, useful therapeutically

AA;

61.4%; Score 3916; DB 2; Length 756;

arity 94.9%; Pred. No. 3.9e-271; Indels 36; Gaps 1;

conservative 0; Mismatches 4; Indels 36; Gaps 1;

SQSQGDPLPPAHTGYRTSTAPSSPGVDSVPLQRTGSHQPNAAAATFORASYAA 493

SQSQGVPLPPAHTGYRTSTAPSSPGVDSVPLQRTGSHQPNAAAATFORASYAA 60

INADPYRQLOPCPSVESYSKSGPALPEGLTARSIDSIOKOPREGWRDPELP 553

INADPYRQLOPCPSVESYSKSGPALPEGLTARSIDSIOKOPREGWRDPELP 120

MLQHOFPSVQSNAAAYLQHLCEGDNKIKAERROGGIQLLVLDLDRHTEVRSAC 613

MLQHOFPSVQSNAAAYLQHLCEGDNKIKAERROGGIQLLVLDLDRHTEVRSAC 180

NLVYGKANDDNKIALKNCGGIPALVRLRKTDLLEIRELVGTGLVNLSSCDALKMP 673

NLVYGKANDDNKIALKNCGGIPALVRLRKTDLLEIRELVGTGLVNLSSCDALKMP 240

ALAVLTNAVIIIPHSGWENSPLODRKIQLHSSQVLRNATGCLRNYSAGEARRM 733

ALAVLTNAVIIIPHSGWENSPLODRKIQLHSSQVLRNATGCLRNYSAGEARRM 300

GLTLDALLVVIQALGSSEIDSKVENCVCILRNLSVRLAAETSOQOHMTDDELGL 793

GLTLDALLVVIQALGSSEIDSKVENCVCILRNLSVRLAAETSOQOHMTDDELGL 360

ANGKDAESGCGWKKKKKKKQDQWDGVGPLPDCAEPKGIQMLWHPISVVKPYLT 853

ANGKDAESGCGWKKKKKKKQDQWDGVGPLPDCAEPKGIQMLWHPISVVKPYLT 420

SNPDTLEGAGALQNLAAAGSWKSVYIRAAVRKEKGLPLVLLRIDLNDRVACAVA 913

SNPDTLEGAGALQNLAAAGSWKSVYIRAAVRKEKGLPLVLLRIDLNDRVACAVA 480

RNALDVNRKELLIGKAMRDLVRLPGGNNNNNTASKAMSDDTVTAVCCTLHEVITK 973

RNALDVNRKELLIGKAMRDLVRLPGGNNNNNTASKAMSDDTVTAVCCTLHEVITK 540

YAKALRDAGGIEKLVIKSGKDKHSPKVVKAASQVLNSMWQYRDLRSLYKKGWSQ 1033

YAKALRDAGGIEKLVIKSGKDKHSPKVVKAASQVLNSMWQYRDLRSLYKKGWSQ 600

IASSTIERDRQPYSSSRTPSISPVRSVNNNSASAPASPREMISUKERKTDYECT 1093

IASSTIERDRQPYSSSRTPSISPVRSVNNNSASAPASPREMISUKERKTDYECT 660

ATHGKAGETSCKDAMTAQNTGISTLYRNSYCAPAEDIKHQVSAQVPQEPSRKD 1153

ATHGKAGETSCKDAMTAQNTGISTLYRNSYCAPAEDIKHQVSAQVPQEPSRKD 692

IQPFQNSTRYNDSFFEDQVHRPPASEYTMHLGKSTGNYVDYFYAARPYSEINYE 1213

-----NSTRYNDSFFEDQVHRPPASEYTMHLGKSTGNYVDYFYAARPYSEINYE 744

YPASPDSSWV 1225

YPASPDSSWV 756

ndard; protein; 1192 AA.

AC AAY23899;

XX 27-SEP-1999 (first entry)

XX Human resenilin binding armadillo protein p0071.

DE Human resenilin binding armadillo protein p0071.

XX Human; presenilin 1; PS1; presenilin-binding protein; interactin

KW presenilin allele; Alzheimer's disease; senile dementia;

KW psychiatric disease; schizophrenia; depression; neurological dis

stroke; cerebral haemorrhage; p0071; armadillo protein.

XX Homo sapiens.

OS WO9935501-A1.

XX 15-JUL-1999.

XX 08-JAN-1999; 99WO-CA0000018.

XX 09-JAN-1998; 98US-0070948P.

(UTOR ) UNIV TORONTO GOVERNING COUNCIL.

XX St George- Hyslop PH, Fraser PE;

XX WPI; 1999-419410/35.

XX Identifying substances that alter presenilin interactions, usefu

PT screening individuals for presenilin alleles associated with Alz

PT disease - useful for diagnosis of Alzheimer's disease.

XX Claim 10; Fig 1; 40pp; English.

XX The present sequence represents a human presenilin-binding prote

CC termed p0071. It is a member of the armadillo family of proteins

CC the interaction of a presenilin with a presenilin-binding protei

CC method comprises contacting the interacting domain of a presenil

CC protein to a presenilin-binding protein in the presence of a tes

CC substance, and measuring the interaction of the presenilin and t

CC for presenilin alleles associated with Alzheimer's disease and r

CC disorders, such as senile dementia's, psychiatric diseases such

CC schizophrenia and depression, and neurological disease, such as

CC and cerebral haemorrhage

XX Sequence 1192 AA;

SQ

Query Match 44.2%; Score 2818.5; DB 2; Length 1192;

Best Local Similarity 49.2%; Pred. No. 3e-192;

Matches 618; Conservative 192; Mismatches 337; Indels 109;

QY 15 MPVDPQSSASEKTSLSPLNTSNGDSETETT-SAILASVKEQELQERLTRE

Db 1 MPAEQASLVEGQPTQREA-ASTGPGNEPETTATILLASVKEQELQERLTRE

QY 74 QIVASQLERCKLSETGSMSSSAEPOFQWQSD-----QKQIEDLTTGLE

Db 60 QIVASQLERCKLGAESPSIASTSTSEKSPWRSTDVNTGSKPRVSDAVQPNNY

QY 128 IRSLSQSGIL-PPDYSTGERFSL--SQSALQASKPEGSFYPAHYHNTLA

Db 116 IRTEPEQGLYSPQOTSLHESGSLGNSRSTOMNSYSDSGYQAGSFNSQNS

QY 180 -----LGETTPSQLPARGTQARATQSFQ-GTTSRAGHLAGEPAPPPPP----

Db 176 QQHSFICSTNNHV-----RNSRAEGTLLVQPSVANPAMERVSVSPRAQSPYVI

QY 225 PPREPPAPSLGSAFHLP----DAPPAALAAALYSSTLPPAPRGSGPLAAPQGS

Db 232 PSRGLRSLTSLGSGFGSPSVTDPRPLNPSA--YSSTLLPA-ARAASPY-SQRPAS

QY 282 RGSAPGATYAAPRGSSPK-QSPSLAKSYSTSSPINIVSSAGLSPIRVTSP

andard; protein; 1211 AA.

interacting protein p0071.

XX		Presenilin-interacting protein; human; Alzheimer's disease; dia
KW	KW	therapy; transgenic animal; animal model; p0071.
XX	OS	Homo sapiens.
XX		
XX	Key	Location/Qualifiers
PH	Domain	509..1022
FT		/note= "presenilin-interacting domain"
XX	WO9727296-A1.	
PN		
PD	31-JUL-1997.	
XX		
XX	27-JAN-1997;	97WO-C A000051.
PF		
XX	26-JAN-1996;	96US-00592541.
PR	05-JUL-1996;	96US-0021673P.
PR	12-JUL-1996;	96US-0021700P.
PR	08-NOV-1996;	96US-0029895P.
PR	02-JAN-1997;	97US-0034590P.
XX		
PA	(HSCR-) HSC RES & DEV LP.	
PA	(UTOR ) UNIV TORONTO GOVERNING COUNCIL.	
XX		
PI	St George- Hyslop PH, Fraser PE, Rommens JM;	
DR	WPI; 1997-393684/36.	
DR	N-PSDB; AAAT79968.	
XX		
PT	Presenilin-interacting protein genes - used to develop products	
PT	diagnosis, therapy and study of Alzheimer's disease and related	
PT	disorders.	
XX		
XX	Claim 1; Page 104-107; 133pp; English.	
XX		
CC	p0071, a human protein with 'armadillo' repeats, has been ident	
CC	presenilin-interacting protein (PIP) . A yeast two-hybrid kit wa	
CC	cc to screen a human brain cDNA library for clones which interacte	
CC	presenilin transmembrane 6-7 loop domain; mutations in this loo	
CC	are known to be causative of Alzheimer's disease (AD). 9 PIP ge	
CC	sequences (see AAAT79966-74) including p0071 were identified. PI	
CC	acids, PIP proteins and peptides (especially the presenilin int	
CC	domain), antibodies to PIPs, cells transformed with PIP nucleic	
CC	and transgenic animals altered with PIP nucleic acids can be us	
CC	diagnosis, therapy and study of AD and related disorders. They	
CC	used to identify compounds which can modulate the expression of	
CC	gene or which bind to a PIP or modulate its activity	
XX		
SQ	Sequence 1211 AA;	
Query Match	44.2%; Score 2818.5; DB 2; Length 1211;	
Best Local Similarity	49.2%; Pred. No. 3.1e-192;	
Matches	618; Conservative 192; Mismatches 337; Indels 109;	
QY	15 MPVPDPSSASEKTSLSPLNTNSGDSGETTT-SAILASVKOELOFERLTR	
Dd	1 MPAPEQASLVEEGQPOTROEA-ASTGPGMEPPETTAITLASVKOELOFQLTR	
QY	74 QIVASQLERCKLGSETGTSMSSSAEEQFWQSOD-----GQKDIEDELTTGL	
Dd	60 QIVASQLERCRLGAESPSTASTSSTEKSFPWRSTDVNTGVSKPRVSDAVQPNN	
QY	128 IRLSQESGIL-PDQYSTGERSELL--SQSALOLNSKEGFSQYPASTVSHNQTL	
Dd	116 IRTEPQGTYLPFEQTSLHESEGGSLNGRSSTQMNSYSDSGYQEAGSPFNQNV.	
QY	180 ----LGETTTPSQLPARGTQAATGSFSO-GTTSRAGHLAGEPAPPPP----	
Dd	176 QQHSFFTGSTNNHV-----RNSRAEQGLTVQPSVANRRRVSSVPSRAQSFYV	
QY	225 PPREPPAPSLGSAFHLP----DAPAAAAAALYSSTLTLPAPPRGSGSLAAPQGG:	

SLRTSLGSGFGSPVTDPRPLNPSA---YSSITLPA-ARASPY-SQRPASPTAIR 286  
 APEGATVAAPRGSGPK-QSPRLAKSYSTSSPINIVSSAGLSPIRVTSPTVQST 340  
 VTRQT-SNPNPGPTPOQTARV-----GSPILTDAQT-----RVASPSQGO-V 333  
 PIHQLSSTIGTYATLSPT-KRLVHASQYKSHSOELYATATLORPGSLAAGSRASY 399  
 P--KRSQMTAVPQHLGFLQRTVHDMQFGQQQYDIYERMPRPDSL-TGLRSSY 390  
 ZHIGPELALQSPHHIDPIYEDRVYQKPMRSLSQSGDPLPAHTGTTRTSTAP 459  
 SOLGQDLRSVSPDLHTPTIEGRTYSPYRSPNHGTV-LOGSQTALYRT---- 445  
 VDSV-PLQRTGSHQGPCNAATAAATQFASVAGPASNYADPVROLQYCPSPVSK 518  
 VSGIGNLQRTSSQR-----STLTQRYNNYALNTTATYAEPIQYRVQ-ECYNR 496  
 LPPEGTLARSIDSIOKDPREFQWRDPELPEVIQMLQHQFPSPVQSNAAAYLQHL 578  
 VPADDGTRSPSIDSIOKDPREFQWRDPELPEVIHMLHQFPSPVQSNAAAYLQHL 556  
 KIXABIRQGGIQLLVLLDHRMTEVRSAGCALRNLYGKANDDNKIALKNCGGI 638  
 KVMVEVCLGGIKHLVLLDHRVLEVKNAAGCALRNLYGKANDDNKIALKNCGGI 616  
 RLLRKTDLREIRELVTVGLVNLSCDALKWPIIQDALAVLTNAVIIIPHSWENSPL 698  
 RLLRKTDLREIRELVTVGLVNLSCDALKWPIIQDALAVLTNAVIIIPHSWENSPL 676  
 KIQHSSQVLRNATGCLRNVSAGAEARRMRECDGLTDLALYVQISALSGSEIDS 758  
 KIKFQTSVLRLNTGCLRNLTSAAGEARKQWRSCEGLVDSLLYVHCNTSDYDS 736  
 NCVCILNLSVRLAAETSQOHMGCTDELGLLCEANGKDAESGCGWKKKKKK- 817  
 NCVCILNLSVRLAAETSQOHMGCTDELGLLCEANGKDAESGCGWKKKKKK- 795  
 DQWDGVPPLDCABPPKGIQMLHPSIVKPYLTLLSCSPDTEGAGALQNLAA 875  
 DQWDGVPPLDCABPPKGIQMLHPSIVKPYLTLLSCSPDTEGAGALQNLAA 855  
 WSVYIRAAVKEKGLPILVELLRIDNRVVCATARNALDVRNKEKLGKVMR 935  
 FAAYIRGRPKRKGILPILVELLRIDNRVVCATARNALDVRNKEKLGKVMR 915  
 RLPGGNNNTASKAMSDDTVTAVCCTLHEVITKNMENAKALRDAGIEKLVGSK 995  
 RLPGGNGPS-----VLSDETMAALCCALHEVITKNMENAKALRDAGIEKLVGSK 970  
 KHSPKVYKAAQVNLNSMWQYRDLRLSKYKQWQSYHFVASSSTLTERDRQRPYSRR 1055  
 RSSLKVYKAAQVNLNSMWQYRDLRLSKYKQWQSYHFVASSSTLTERDRQRPYSRR 1026  
 SPV--RVSPNNRSASAPSPREMTSLKERTDYCTGTSNATYHGAKEHTSRKDM 1113  
 STTNQCMSPITQSGVSTSSPALLGIRDPRSEYDRTQPMQYNSQGDATHK--GL 1083  
 TGISTLYRNSYGAPED-----IKENQVSAQVPQEPSRKQVETYPQCNSTRNDE 1169  
 SKSPFIYISSYSSPANNRRLQHQQLYYS--QDSSNRKNPDVRLYLQSPHSYED 1141  
 DQVHRRPPASEYTHLGLKSTGNTVDYFSAARPYSELYNETSHYPASPDSWV 1225  
 DRVHF-PASTDYSTQYGLKSTNTYVDFYVSTKRP---SYRAEQYPGSPDSWV 1192

dard, protein; 1211 AA.

21-AUG-2000 (first entry)  
 Human p0071.  
 NIK1 Interacting Protein; p0071; protein complex; cytostatic; ant  
 neuroprotective; cardiant.  
 Homo sapiens.  
 W0200020448-A2.  
 13-APR-2000.  
 06-OCT-1999; 99WO-US023314.  
 06-OCT-1998; 98US-00167206.  
 (CURA-) CURAGEN CORP.  
 Nandabalan K, Schulz VP, Yang M;  
 WPI; 2000-303742/26.  
 N-PSDB; AAA09308.  
 New complex of a NIK1 protein and a NIK1 protein-interacting prot  
 useful for treating cancer, hyperproliferative disorder,  
 neurodegenerative disorder, cardiomyopathies, viral infections ar  
 metabolic disorders.  
 Example 1; Page 154-159; 172pp; English.  
 AAY92331-37 were isolated in a modified yeast two hybrid system v  
 NIK1 protein as "bait". These are known sequences which are NIK1  
 interacting proteins. The invention concerns purified complexes c  
 protein and a NIK1 protein-interacting protein, where the interac  
 protein is chosen from trkA, protein phosphatase lalpa, 14-3-3ef  
 alpha-tropomyosin, vimentin, p0071, Inl-1, IP-1, IP-2, IP-3, IP-4  
 5. NIK1 (also referred to as Nek2) is a human homologue of the  
 Aspergillus nidulans mitotic regulator, NIMA kinase. NIK1 is a  
 serine/threonine-specific kinase and is thought to play a key rol  
 cell-cycle events leading to the onset of mitosis. The complexes,  
 derivatives and NIK1 or NIK1-IP protein and DNA sequences, etc. a  
 useful for treating or preventing a disease or disorder involving  
 aberrant levels of the complex or protein. Such disorders include  
 hyperproliferative disorders, neurodegenerative disorders,  
 cardiomyopathies, viral infections and metabolic disorders  
 Sequence 1211 AA;  
 Query Match 44.2%; Score 2818.5; DB 3; Length 1211;  
 Best Local Similarity 49.2%; Pred. No. 3.1e-192;  
 Matches 618; Conservative 192; Mismatches 337; Indels 109; C  
 QY 15 MPVPDQPSASEKTSKSLSPGLNTSNGDSETETT-SAILASVKEQELQFELTREI  
 DB 1 MPAPEQASLVEEQFQTRQEA-ASTGCMPEPTTATTILASVKEQELQFELTREI  
 QY 74 QIVASQLERCKLGSSTGSMSSSAEQFQWQSQD-----GQKDIDELTTGLEI  
 DB 60 QIVASQLERCKLGSSTGSMSSSAEQFQWQSQD-----GQKDIDELTTGLEI  
 QY 128 IRSIQESGIL-DPODYGTGERPSLL--SQSALQNSKPEGSFOYPASVHSQTLA-  
 DB 116 IRTEPEQGLYSPEQTSLSHESEGLSGNSRSTQMSYSDSGYQEAGSFHNSQNSK  
 QY 180 -----LQETTPSQLPARGTQARATQSFQ-GTTSRAGHLAGPAPPPPP-----  
 DB 176 QQHSFIGSTNNHV-----RNSRAGQTLVQPSVANRWRVSSVPSRAQSPSYIS  
 QY 225 PPREFPAPSLGSAFHLPP---DAPPAALAAALYSSSTLPAPPPRGSGSLAAPQGGSE  
 DB 232 PSRGLRTSLGSGFGSPVTDPRPLNPSA---YSSITLPA-ARASPY-SQRPASE

standard; protein; 1211 AA.

Human p007i protein sequence SEQ ID NO:4487.  
Human; colon cancer; cancer; tissue profiling; forensic; mapping genetic analysis; diagnostic; antisense therapy.

Homo sapiens.

WO200229086-A2.

11-APR-2002.

02-OCT-2001; 2001WO-US030732.

02-OCT-2000; 2000US-0237271P.

(FARB) BAYER CORP.

Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Moli:  
Thiagalingam A, Lewis ME;

WPI; 2002-426115/45.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

new isolated nucleic acid that is differentially expressed in tissues useful for determining the presence of colon cancer in tissue type, and in antisense therapy.

Claim 5; Fig 3; 796pp; English.

ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differ expressed in cancer tissues. ABQ78993 to ABQ79004 represent pro encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) used in antisense therapy. An antibody immunoreactive with a po encoded by (I) is useful for detecting cancer in a patient samp for detecting the presence or absence of a polynucleotide encod nucleic acid which hybridizes to (I) in a cell. A probe/primer from (I) can be used for determining the presence of a nucleic acid hybridizes to (I), and for determining the phenotype of cells in colon cancer in a patient. (I) is useful for determining the pres of cells from a patient. (I) is useful for determining the pres state of other type of cancer, in antisense therapy, to generat macroarrays on a solid surface, to identify a chromosome on whi corresponding gene resides, and in tissue profiling, forensics, analysis, mapping and diagnostic applications. (I) can be used antibodies, and to screen for peptide analogues and antagonists

Sequence 1211 AA;

Query Match 44.2%; Score 2818.5; DB 5; Length 1211;  
Best Local Similarity 49.2%; Pred.No. 3.1e-192;  
Matches 618; Conservative 192; Mismatches 337; Indels 109;

QY	15	MPVPQPSASBKTSLSLPGLNNTSGDGSFTTT-SAILASVKQOELOFRELTRI
Db	1	MPAPQASLVEEQGPQTRQEA-ASTGPMPEPTATTILASVKQOELOFQRLTRI
QY	74	QIVASQLERCKLGSFTGSMSSMAEBOFQCSQD-----GQKDIDELTGLI
Db	60	QIVASQLERCLGASPSIASTSTKSFPMRSIDVPNTGSKPRVSDAVOPNN

```
Qy      128 IRSLQESGIL-DPQDYSTGERPSLL--SQSALQLNKSKEPGSFQYPASYHSNTGLA
       ||| : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      116 IRTEPEQGTLYSPEOTSLHEGSGLSNGSRSSSTOMNSYSDSGYOAGSGFHNQNVE
```

QY 180 -----LGETTPSQIPARGTQARATGQSFSQ-GTTSRAGHLAGEPAPPPPP-----  
          : | | :         : | | | :         | | | :         | | | :  
Db 176 QOHSFIGSTNNHV-----RNSRAEGOTLVOPSVANPAMRRVSSVPESRAOSPSYVI

**Qy** 225 PPREFAPSLGSAFLP--DAPPAAAAALYSSSTLPAPPRCGSPLAAPQGGS  
||| ||| | : ||| : ||| : ||| :  
**D6** 232 PSRGLSTSLGGCGSPSVTPRP LNPSA---YSSTLPA-ARAASPY-SORPAS

QY 282 RGGSAPEGATYAAPRGSSPK-QSPRLAKSYSTSSPINIIVSSAGLSPIRVTSPF

VTSTQI-SNPNGPPQYQTTARV-----GSPILTLTDAQT-----RVASPSQGV-V 333  
 PIHQLSSTIGTYATILSPT-KRLVHASBOYSKHSQELVATATLQRPGLSAGSRASY 399  
 P--KRSGMTAVPQHLGFSQRTVHMEQFGQQQYDIYERWVPPRPSDL-TGLRSSY 390  
 GHLGPELRALQSPHEHIDPIYEDRYQYQPPMRSLSSQSGDPLPPAHTGTCTYSTAP 459  
 SOLQDLRSVSPDLHTIPTEGRTYISPVYRSPNHTGVE-LQSQTALYIT---- 445  
 VDSV-PLQRTGSHQHPQNAATFORASYAAGPASNYADPYRQYQCPSPVSPYSK 518  
 VSGIGLQRTSSQR-----STLTQYNNYALNTTATYAEPYRPIQYRQV-ECYNR 496  
 LPPSGTLARSPISDSIQKDPREFGWRDPELPEVQMLQHPSPVSQSNAAVLOHLC 578  
 VPADDGTRSPISDSIQKDPREFAWRDPELPEVHMLEHPFVSQVQNAAYLOHLC 556  
 KIKAEIRROGGIQLLVLLDHRMTVEHRSACALRNLYGKANDNKIALKNCGGI 638  
 KVMVEVCLGGIKHLVDLLDHRVLEQVACALRNLYGKSTDENKIAMKNVGGI 616  
 RLLKRTDLEIRELVTVGLWNLSSCDALKMPIIQDALAVLTNAVITPHSGWENSGPL 698  
 RLLKRSIDAEVRELVTGLWNLSSCDVAKMTIIRDALSTLNTVIVPHSGWNNSSF 676  
 KIQHSSQVLTNATGCLRNVSAGAEARMRBEGDGLTDALLYVQSLQSSSEIDS 758  
 KIKFQTSVLVLRNTTCLRNLYSAGEEARQMRSCGLVDSLLYVHTCVNTSDYDS 736  
 NCVCILRNLYSLAAETSGQHMGTDELDGLLCEANGKDAESSGCGWKKKKKK- 817  
 NCVCILRNLYSLYLEVFPQARILGLNELDGLLKESPSKDSPPS-CWCKKKKKKKR 795  
 DQWDGVGLPDCAEPPKGIQMLWHPISYKPYLTLLSECSNPDFTLEGAGALQNLAA 875  
 DQWDGVGPIPLGLSKSPKGVEMLWHPISYKPYLTLLAESNPATLBSAGSLQNLAA 855  
 MSVYIRAAVREKGIPIILVELLRIDNRVCAVATARNMALDVENKELIGKYAMR 935  
 FAAYIRGGRPKRGKGIPIILVELLRIDNRVCAVATARNMALDVENKELIGKYAMR 915  
 RLPGGNNNTASKAMSDDTTAVCTILHEVITKNMENAKALRDAGGTEKLVGISK 995  
 RLPGGNGFS-----VLSDETMAAICCALHEVITKNMENAKALADSGGTEKLVNITK 970  
 KHSKPVVKAASQVLNSMQRDLRLSLYKDGWSQYHFVASSSTIIBDRQRPYSR 1055  
 RSSLKVVKAAQAQVLTNLQYRDLRLSLYKDGWQNHFTTPVSTLERDRFKSH- 1026  
 SPV--RVSPNNRSAPAPREMIKTKERTDYECTGSGNATVHGAKGHTSRKQDAM 1113  
 STTNQOMSPITQSVGTSSTSSPALLGIRPRSEYDQTPPMQYNSQGDATHK--GL 1083  
 TGISTLYRNSYGAPED----IKHNQVSAQPVQPSKDYETYPQFNSTNRNDE 1169  
 SKPSPIYISSYSSPAREQNRRLHQOQLVYS--QDDSNRKNFADYRLYLQSPHSYED 1141  
 IDVHHRPPASEYTHMLGLKSTGNYDYFVSAAAPYSELNLYTSHYSPASDSWV 1225  
 DRVHF-PASTDYSTQYGLKSTNYDYFVSAAAPYSELNLYTSHYSPASDSWV 1192

idard; protein; 1110 AA.

(first entry)

ellular matrix and cell adhesion molecule-12 (XMAD-12).

XX Human; extracellular matrix and cell adhesion molecule; XMAD;  
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;  
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dy;  
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocar;  
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS  
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; me;  
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma;  
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoar;  
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative coliti;  
 KW infection; cell proliferative disorder; actinic keratosis; myeloi  
 KW arteriosclerosis; nontropic; anticonvulsant; antithyroid; nephroi  
 KW neuroprotective; dermatological.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 FH 511..1025 /note= "Mouse p120 protein"  
 FT Region  
 FT 532..565 /note= "Armadillo/beta-catenin-like repeats"  
 FT 577..598 /note= "Armadillo/beta-catenin-like repeats"  
 FT 609..622 /note= "Armadillo/beta-catenin-like repeats"  
 FT 971..973 /note= "Rgd cell interaction motif"  
 FT WO200142285-A2.  
 PD 14-JUN-2001.  
 XX 05-DEC-2000; 2000WO-US032990.  
 XX 10-DEC-1999; 99US-0172852P.  
 PR 16-DEC-1999; 99US-0172354P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;  
 PI Baughn MR, Lu DAM, Shah P, Au-Young J;  
 PI WPI; 2001-381632/40.  
 DR N-PSDB; AAD08056.  
 XX New human extracellular matrix and cell adhesion molecules and  
 PT polynucleotide sequences encoding them, useful for diagnosis, pr  
 FT treatment of genetic, autoimmune and cell proliferative disorder.  
 XX Claim 1; Page 104-107; 135pp; English.  
 CC The present sequence is a human extracellular matrix and cell ad  
 CC molecule (XMAD). The XMAD is used for screening a compound for  
 CC effectiveness as an agonist or antagonist of XMAD. The identifie  
 CC or antagonist are used for treating a disease or condition assoc  
 CC with decreased or increased expression of functional XMAD. The  
 CC polynucleotides encoding XMAD are useful in somatic or germline  
 CC therapy to correct a genetic deficiency, to express a conditiona  
 CC lethal gene product and to express a protein which affords prote  
 CC against intracellular parasites and also for diagnosis of disord  
 CC associated with expression of XMAD. They are also used for gener  
 CC hybridisation probes useful in mapping the naturally occurring g  
 CC sequences and to create knock in humanised animals (pigs) or tra  
 CC animals (mice or rats) to model human diseases. Oligonucleotide  
 CC fragments derived from the polynucleotide sequences may be used  
 CC elements on a microarray. Antibodies which specifically bind XMAD  
 CC used for the diagnosis of disorders associated with the expressi  
 CC XMAD, or in assays to monitor patients being treated with XMAD. I  
 CC diagnosed, prevented or treated include genetic disorders such a  
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher'  
 CC disease, myotonic dytrophy, sickle cell anaemia, thalassaemia,  
 CC autoimmune/inflammatory disorders such as acquired immune defici  
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,

sis, Crohn's disease, diabetes mellitus, atopic dermatitis, arthritis, multiple sclerosis, Grave's disease, osteoarthritis, psoriasis, rheumatoid arthritis, ulcerative colitis, fungal, parasitic, protozoal and helminthic infections and rative disorders such as actinic keratosis, arteriosclerosis including breast, bladder, bone marrow, brain and uterus aemia, adenocarcinoma, lymphoma, melanoma and myeloma

0 AA;

41.8%; Score 2665.5; DB 4; Length 1110;

larity 51.5%; Pred No. 2.5e-181;

Conservative 164; Mismatches 292; Indels 95; Gaps 32;

/PDPASASEKTSLSPLGNTSNGSGSETT-SAILASVKEQELQFELTLRELEAER 73

PEQASLVEEQPQTRQEA-ASTGPGMEPETTATTILASVKEQELQFELTLRELEVER 59

/ASQLERCKLGSFTGSMSSNAEEQFQSQD-----GQKDIEDELTGLLELVDS 127

/ASQLERCKLGAESPASIASTSTSEKSPFMRSTDPVNTGSKPRVDAVQPNYL----- 115

ILQESGIL-DPODYSTGERPSLL--SOSALQLNSKEGSGFOYPASVHNOTLA----- 179

PEQGTLSPEQTSLSHEGSLGNSRSSTQMSYSDSGYQAGSFHNSQVSKADNR 175

---LGTTTSPQPARQTARATQSFQ-GTTSRAGHLAQPFPAPPPPP----- 224

ISPIGSTNNHV-----RNSRAEGOTLVQPSVANRMRVSSVPSRAQSPSYVISTGVS 231

REFPAPSLGSAFLP--DAPPAALAAALYYSSSTLPAPPRGSPLAAPQGSPTKLQ 281

QSLRTSLSGGSGSPVTPRPLNPSA--YSSSTLPA-ARAASPY-SORPASPTAIR 286

ISAPEGATYAAPRGSPK-QSPSLAKSYSTSPINIVVSSAGLSPIRVTSPTVOST 340

ISVTSQGT-SNPNPTPQYQTVARV-----GSPFLTDAQT-----RVASPSQGV 333

ISPIHQLSTIGYATLSPT-KRLVHASQYKSHSQELYATATLQPPGSLAGSRASY 399

ISP--KRSQMTAVFQHLGSLQRTVHDMEQFGQQQYDIYERVMVPPRPDSL-TGLRSSY 390

IGHGLPELRALQSPHHIDPIYEDRVYQKPPMRSLSQSQDPLPPAHTGTVRTSTAP 459

HSQLGQDLRSVSPDLHTPIYEGRTYSPVYRSPNHGTVB-LOGSQALYRTGS-- 447

GVDSVPELQRTSQHPQNAATAFORASYAAGPSNADPYRQLOQCPSPVESPSYSKS 519

-GIGN--LQRTSSQ-----STLTYQRNNVALNTTATYAEPIQYRVQ-ECNYNRL 496

LPPEGTLAARSPISDSIQKDPREFGWRDPPELPEVIOHLOQFPSPVQSNAAVYLHLCP 579

VPADGDTTRSPSIDSIQKDPREFAWRDPPELPEVIMHLOQFPSPVQNAAYLHLCP 556

IKIKABIRQGGTQLLDLDRHMTVHVSACGALRNLYVYGKANDNKIALKNCGGIP 639

IKVMEVCLGGIKHVLVDLDRHVLVQVQACGALRNLYVFGKSTENKIAMKNVGGIP 616

RLLRKTTDLRELVTVGLVNLSSCDALKMPIIODALAVLTNAVIPHSGWENSPLO 699

ALLRKSIDAEVRELVTVGLVNLSSCDALKMPIIODALAVLTNAVIPHSGWENSPLO 676

IKQLHSSQVLRNATCLNRVSSAGEARRRRECGLTDALLYVIOQSALGSETDSK 759

IKIKFTSLVLRNTTCLNRSLSSAGEARRRRECGLTDALLYVIOQSALGSETDSK 736

ENCVCILRLNLYSLAETSOGQRMGTDELGLLCCGANGKDAESSCGWKKKKKK-- 817

ENCVCILRLNLYSLAETSOGQRMGTDELGLLCCGANGKDAESSCGWKKKKKK-- 795

QDQDVGVLPCDAEPKQIMLWHPISVVKPYLTLSECSNPDTLEGAGALQNLAA 876

QDQDVGVPITGLSKSPKGVEMLWHPISVVKPYLTLSECSNPDTLEGAGALQNLAA 855

QY 877 SWKMSVYIRAAVRKEKGLPIILVELLRIDNRVVCVATALRNMALDVNRKELIG

Db 856 NWKFAAYIRAAVRKEKGLPIILVELLRIDNRVVCVATALRNMALDVNRKELIG

QY 937 LVHRLPGNNSNNTASKMSDDITVAVCCYLHEVITKMNENAKALRDAGGIEKL

Db 916 LVNRLPGNGGFS-----VLSDETMAAICCAHEVTSKMNENAKALDGGGIEKL

QY 997 KGDHSPKVVKAASQVLNSMQRDLRSLYKKGQWSQVHFVASSSTIERDRQRP

Db 971 RGDSSLLKVVKAASQVLNTLWQYDLRSYKKGQWNQNHFTPTVSTLERDRFKSI

QY 1057 PSISPV--RVSPNNRSASAPSPREMSLKERKTDYECTGSNATYHGAKGHEHTSI

Db 1026 PSLSTTQQMSPIQSVGSTSSSPALLGIRDPRSEYDRTQPPMYNSQGDATHI

RESULT 12

ABP41836

ID ABP41836 standard; protein; 785 AA.

XX

AC ABP41836;

XX

DT 22-AUG-2002 (first entry)

XX

DE Human ovarian antigen HOCPT34, SEQ ID NO:2968.

XX

KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

KW ovarian cancer; breast cancer; tumour; reproductive system diso;

KW infertility; pregnancy disorder; anovulation; polycystic ovary;

KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infectio;

KW inflammatory condition; immune disorder; blood disorder;

KW cardiovascular disorder; respiratory disorder; neurological dis;

KW gastrointestinal disorder; urinary system disorder; drug screen;

KW gene therapy; chromosome mapping; forensic analysis;

KW antibody preparation; cytostatic; immunomodulatory; neuroprotect

KW antiinflammatory; gynaecological; reproductive; chromosome 2q23.

XX

OS Homo sapiens.

XX

PN WO200200677-A1.

XX

PD 03-JAN-2002.

XX

PF 07-JUN-2001; 2001WO-US018569.

XX

PR 07-JUN-2000; 2000US-0209467P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Birse CE, Rosen CA;

XX

DR WPI; 2002-147878/19.

XX

DR N-PSDB; ABQ54913.

XX

XX Isolated nucleic acid molecules encoding novel ovarian polypept;

PT useful in the prevention, treatment and diagnosis of cancer (e.g;

PT cancer), immune disorders, cardiovascular disorders and neurolog;

PT diseases.

XX

PS Claim 11; SEQ ID NO 2968; 2922pp; English.

XX

XX The invention relates to 2175 novel human ovarian antigens (ABP;

CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and a;

CC encompasses polypeptides 90% identical and polynucleotides 95%;

CC to the sequences of the invention. The invention additionally re;

CC recombinant vectors and host cells comprising human ovarian anti;

CC polynucleotides, antibodies against human ovarian antigens, and

CC of ovarian antigen polynucleotides and polypeptides in diagnosi;

CC treating, prognosing or preventing various ovary and/or breast-

CC disorders. Such conditions include ovarian cancer and breast ca;

CC metastatic tumours of ovarian or breast origin, reproductive sy;

```

QY      1046 DRQPYSSRTPSPV--RVSPNNRSASAPASPREMISLKERKTDYECTGSNAT
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB      705 DRFKSH-----PSLSITTNQMSPIIQSVGSTSSPALLGIRDPRSEYDRTPPMQ
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY      1104 GEHTSR 1109
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB      760 GDATHK 765

RESULT 13
ABG04995
ID ABG04995 standard; protein; 696 AA.
XX
XX
AC ABG04995;
XX
XX      13-FEB-2002 (first entry)
DT
DE Novel human diagnostic protein #4986.
```

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX  
XX  
TOO SUPPLEMENT

OS Homo sapiens.

XXIX

PN WO200175067-A2.  
XX  
PD 11-OCT-2001.

PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US0008631.

PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.

PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX

XX  
PA (HYSE-) HYSEQ INC.  
XX

XX PI Drmanac RT, Liu C, Tang YT;  
XX

XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS69182.

DR N-PSDB; AAS69182.  
XX  
PT New isolated polynucleotide a

PT New isolated polynucleotide a  
PT diagnostics, forensics, gene  
PT responsible for genetic disorder

PT responsible for genetic disorder  
PT biodiversity.  
XX

XX  
PS Claim 20; SEQ ID NO 35354; 10  
XX

XX The invention relates to isol  
CC sequences. (I) is useful as h  
CC

sequences. (I) is useful as a  
reaction (PCR) primers, oligo  
and in recombinant production

CC and in recombinant production  
CC in diagnostics as expressed s  
CC genes. (I) is useful in gene

genes. (I) is useful in gene activity of (II) or to treat useful for generating antibody

useful for generating antibodies in tissue, as mol  
polypeptide in tissue, as mol  
supplement. (II) and its bind

CC supplement. (II) and its bind  
CC of sites expressing (II). (I)  
CC involving aberrant protein ex

CC involving aberrant protein ex  
CC polypeptide and polynucleotid  
CC diagnostics, forensics, gene

CC diagnostics, forensics, gene  
CC responsible for genetic disorder  
CC and to produce other types of

CC and to produce other types of  
CC amino acid sequences. ABG0001  
CC amino acid sequences of the i

CC amino acid sequences of the 1  
CC patent did not appear in the  
CC electronic format directly fr

CC electronic format directly from  
CC ftp.wipo.int/pub/published\_pc  
XX

XX Sequence 696 AA;

Query Match 35.9%;

larity 78.9%; Pred. No. 9.2e-155;  
 Conservative 13; Mismatches 19; Indels 90; Gaps 3;  
 EQFQWSQDQKDIEDELTTGLELVDSCIRSLQESGILDDPQYSTGER-----147  
 EQFQWSQDQKDIEDELTTGLELVDSCIRSLQESGILDDPQYSTGERVIHFGKTLG120  
 -----147  
 EDYSIMEXQMVKVETVTPTQVSVLKANETMASALCGQEGHDDPLLLTVQMSTGNHYI180  
 -----PSLLSQSALQNLNSPEGSFQYPASYHNSQTLALGETTPSQLPARGTOAR196  
 ISLFPWKVPSLLSQSALQNLNSPEGSFQYPASYHNSQTLALGETTPSQLPARGTOAR240  
 IQSFSQGTTSRAGHLACPEPAPPPPPPPPPPPAPSLGSAFHLDPAPPAALYYSS256  
 IQSFSQGTTSRAGHLACPEPAPPPPPPPPPPPAPSLGSAFHLDPAPPAALYYSS300  
 JAPPPRGGSLAAPQGGSPTKLQRGGSAPGATYAAPRGSSPKQSPSLAKSYSTSSP316  
 JAPPPRGGSLAAPQGGSPTKLQRGGSAPGATYAAPRGSSPKQSPSLAKSYSTSSP360  
 TVSSAGLSPIRVTSPTTVQSTTSSSPIHQLSSTIGTYATLSPTKELVHASEQYSKHS376  
 TVSSAGLSPIRVTSPTTVQSTTSSSPIHQLSSTIGTYATLSPTKELVHASEQYSKHS420  
 JXATATLQRPGLAAGSRASYSSQHGHLGPELRALQSPHHIDPIYEDRVYQKPPMRS436  
 JXATATLQRPGLAAGSRASYSSQHGHLGPELRALQSPHHIDPIYEDRVYQKPPMRS480  
 JSQGDPLPPAHTGTYRTSTAPSSPGVDSVPLQRTGSHQGFQNAANAATFORASYAAGPA496  
 JSQGDPLPPAHTGTYRTSTAPSSPGVDSVPLQRTGSHQGFQNAANAATFORASYAAGPA540  
 ADPYRQLQYCPVESPYSKSGPALPPEGTLARSPSIDSIQKDPREF-----GWRDPEL552  
 ADPYRQLQYCPVESPYSKSGPALPPEGTLARSPSIDSIQKDPKYPILCDGNGNPF600  
 TQMQLQHQPFSVQSNAAAYLQHLFCFGDNKIKAEIRR590  
 TQMR-----PSSRRERASSLLKSMCYQXNRILYKFER632  
 undard; protein; 1009 AA.  
 (first entry)  
 diagnostic protein #20811.  
 nosome mapping; gene mapping; gene therapy; forensic;  
 ment; medical imaging; diagnostic; genetic disorder.  
 3.  
 -A2.  
 ; 2001WO-US008631.  
 ; 2000US-00540217.  
 ; 2000US-00649167.  
 3Q INC.  
 Liu C, Tang YT;  
 9362/73.





ILTEAYEFATSVLAGSSVSTLPHLLAPKLIHAWSLAERGRKSEAQQYCDIAAAL 1311  
KXKG-----DKHSPKVVKAASQVLNMQYRDLRLSLYKKG--WSQVH-FV 1037  
KPSGYHNOHLFFGVDELARLRETTSDG-GSSWISRP--SWEKVSGSMMAKFNFSV 1368  
---ASSSTIERDRQRPYSS-----SRTPSIS-----PVRVS 1065  
3DAASTGSKVAEBIGPFARYSGTPTISRSPSVSDIYGSYPVAAAQPLPATGPSRYQ 1428  
3ASAPASPREMISLKERKTDYECTGNSATYHGAKGHTSRKDAMTAQNTGI----- 1120  
ZAPS-ASPEQL-----RGRSSMDSQRSASFYPLGQRGSOEESTPYDTNMFHGM 1482  
-----STLYRNSY-----GAPAEIK-HNOVSAQVPQEPSRKDYET 1156  
?PVAGYQSTPPQSSYMLAPVAEDSASGAQOESFSAHSQVS-----DNAPSHRS-ST 1537  
?FQN--STRNYDESFEDQVHHRPPAS 1183  
?FGHPEDTQAVSTTSQPDQGGYMPPTS 1568

April 22, 2004, 12:21:03

14:26:18 2004

us-09-501-171a-4.rai

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

in search, using sw model

```

cil 22, 2004, 12:11:45 ; Search time 23 Seconds
                          (without alignments)
                          2749.644 Million cell updates/sec

```

-09-501-171A-4  
77  
MFARKPPGAAPLGAMPVDO.....PKSELNVTSTHYPASPSGW 1225

CSUM62

cop 10.0 , Gapext 0.5

9414 seqs, 51625971 residues

389414

zth: 0

00000000007 :H3E

inimum Match 0%

aximum Match 100%  
isting first 45 summaries

```

issued Patents AA: *
/cgn2_6/ptdata/2/iaa/5A_COMB.pcp:*
/cgn2_6/ptdata/2/iaa/5B_COMB.pcp:*
/cgn2_6/ptdata/2/iaa/6A_COMB.pcp:*
/cgn2_6/ptdata/2/iaa/6B_COMB.pcp:*
/cgn2_6/ptdata/2/iaa/PTUS_COMB.pcp:*
/cgn2_6/ptdata/2/iaa/backfiles.pcp:*

```

the number of results predicted by chance to have a  $r$  than or equal to the score of the result being printed, as determined by analysis of the total score distribution.

## SUMMARIES

Entry	Each			ID	Description
	Length	DB			
3.1.2	1084	4	US-09-227-725A-3	Sequence 3, Appli	
3.1.4	756	3	US-08-982-785A-2	Sequence 2, Appli	
3.1.5	686	3	US-08-982-785A-8	Sequence 8, Appli	
3.1.7	620	3	US-08-982-785A-10	Sequence 10, Appli	
3.1.12	1211	4	US-09-167-206-14	Sequence 14, Appli	
3.1.13	1193	4	US-09-227-725A-4	Sequence 4, Appli	
3.1.9	682	3	US-08-982-785A-9	Sequence 9, Appli	
3.1.10	666	3	US-08-982-785A-11	Sequence 11, Appli	
3.1.11	1195	3	US-09-041-886-23	Sequence 23, Appli	
3.1.3	1184	4	US-09-266-225D-18	Sequence 18, Appli	
3.1.6	44	3	US-08-982-785A-4	Sequence 4, Appli	
3.1.2	39	3	US-08-982-785A-6	Sequence 6, Appli	
3.1.2	41	3	US-08-982-785A-5	Sequence 5, Appli	
3.1.1	1601	4	US-09-345-473E-40	Sequence 40, Appli	
3.1.1	1088	4	US-09-233-857F-13	Sequence 13, Appli	
3.1.1	2842	1	US-07-741-940-7	Sequence 7, Appli	
3.1.1	2842	1	US-08-289-548A-7	Sequence 7, Appli	
3.1.1	2842	1	US-08-452-854-7	Sequence 7, Appli	
3.1.1	2842	4	US-08-449-731-7	Sequence 7, Appli	
3.1.0	2843	1	US-08-452-855B-2	Sequence 2, Appli	
3.1.0	2843	1	US-08-452-855B-7	Sequence 7, Appli	
3.1.0	2843	3	US-08-450-582-2	Sequence 2, Appli	
3.1.0	2843	3	US-08-450-582-7	Sequence 7, Appli	
3.1.0	2973	2	US-08-821-355A-7	Sequence 7, Appli	
3.1.0	2973	2	US-09-003-687A-7	Sequence 7, Appli	
3.1.0	2973	3	US-09-136-605-7	Sequence 7, Appli	
3.1.0	41	3	US-08-982-785A-7	Sequence 7, Appli	

Sequence 1  
Sequence 2  
Sequence 3  
Sequence 4  
Sequence 1  
Sequence 2  
Sequence 3  
Sequence 4  
Sequence 1  
Sequence 2  
Sequence 3  
Sequence 4  
Sequence 1  
Sequence 2  
Sequence 3  
Sequence 4  
Sequence 1  
Sequence 2  
Sequence 3  
Sequence 4

## ALIGNMENTS

```

RESULT 1
US-09-227-725A-3
; Sequence 3, Application US/09227725A
; Patent No. 6383758
; GENERAL INFORMATION:
; APPLICANT: St. George-Hyslop, Peter H.
; APPLICANT: Rommens, Johanna
; APPLICANT: Fraser, Paul E.
; TITLE OF INVENTION: Alzheimer's Related Proteins and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1034/1F810-US1
; CURRENT APPLICATION NUMBER: US/09/227,725A
; CURRENT FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (115)
; OTHER INFORMATION: Xaa can be any amino acid
US-09-227-725A-3

```

Query Match	88.2%;	Score 5625;	DB 4;	Length 1084;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 1075;	Conservative	0;	Mismatches 3;	Indels 0; G
QY	148	PSLSQSALQINSKPGSFQYPASVHSNQTLALGETTTSQLPARGTQARATGQSFS		
DB	7	PSLSQSALQINSKPGSFQYPASVHSNQTLALETTTSQLPARGTQARATGQSFS		
QY	208	SRAGHLAGPPAPPPPPPPPPFAPSLGSAFHLDPAPAAAAAALYYSSTSLPAPP		
DB	67	SRAGHLAGPPAPPPPPPPPPFAPSLGSAFHLDPAPAAAAAALYYSXTLPAPP		
QY	268	PLAAPQGGSTPKLQRGSAPEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVVSS		
DB	127	PLAAPQGGSTPKLQRGSAPEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVVSS		
QY	328	PIRVTSPTVQSTIISSPHIQLSSITIGIYATLSPTKRLVHASEQYKHSQELYATA		
DB	187	PIRVTSPTVQSTIISSPHIQLSSITIGIYATLSPTKRLVHASEQYKHSQELYATA		
QY	388	PGSLAAGSRASYSSQGHGLGPELRALQSPHHIPIPIVEDRVYQKPPMRSLSQSQGD		
DB	247	PGSLAAGSRASYSSQGHGLGPELRALQSPHHIPIPIVEDRVYQKPPMRSLSQSQGD		
QY	448	AHTQTYRTSTAPSPGVDVSVPLQRTGSGHGPNAAAATFORASYYAAGPASNAYADPY		

```

|||||
GTYTSTAPSPGVDSVPLQRTSGHQFNAAAATFQASAAAGPASYADPYQLQ 366
SVESPYSGPALPEGTGLARSPSIDSKQDPREFGWRDPELPEVIQMLQHFPSVQ 567
SVESPYSGPALPEGTGLARSPSIDSKQDPREFGWRDPELPEVIQMLQHFPSVQ 426
AAYLQHLCFGNKIKAEIRROGGTQLLDVLDLHRMTEVHRSAACALRLVYTGKANDD 627
AAYLQHLCFGNKIKAEIRROGGTQLLDVLDLHRMTEVHRSAACALRLVYTGKANDD 486
ALKNCGGIPALVRLRLKTTDLLEIRELVTVGLVNLSSCDALKMPILQDALAVLTNAVI 687
ALKNCGGIPALVRLRLKTTDLLEIRELVTVGLVNLSSCDALKMPILQDALAVLTNAVI 546
ISGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEEARRMRRECDGLTDALLYVI 747
ISGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEEARRMRRECDGLTDALLYVI 606
LGSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTDLDGLLCEANGKDAESSG 807
LGSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTDLDGLLCEANGKDAESSG 666
KKKKKKKSQDQDVGVPDCAEPPKGIQMLWHPISVVKPYLTLLSECSNPDTLGGAA 867
KKKKKKKSQDQDVGVPDCAEPPKGIQMLWHPISVVKPYLTLLSECSNPDTLGGAA 726
QNLAAAGSKWKSQVYIRAAVRKEKGLPILVELLRIDNDRVVCATARNMALDVRENKE 927
QNLAAAGSKWKSQVYIRAAVRKEKGLPILVELLRIDNDRVVCATARNMALDVRENKE 786
KYAMRDLVHRLPGGNNNNNTASKAMSDDTVTAVCCTLHEVITKMNENAKALRDAGGI 987
KYAMRDLVHRLPGGNNNNNTASKAMSDDTVTAVCCTLHEVITKMNENAKALRDAGGI 846
VGISKSGDKHS PKVVKAAASQVLSNMWQYRDLRLSLYKKGWSQVHFVASSSTIERDR 1047
VGISKSGDKHS PKVVKAAASQVLSNMWQYRDLRLSLYKKGWSQVHFVASSSTIERDR 906
YSSRTFSPISPRVSPNNRSASAPSPREMISLKERKTDYECTGNSNATYHGAKGEHT 1107
YSSRTFSPISPRVSPNNRSASAPSPREMISLKERKTDYECTGNSNATYHGAKGEHT 966
DAMTAQNTGISTLRNSYGAPAEIDIKHNOVSAQVPQEPSPKDYETYPFQFONSTRNY 1167
DAMTAQNTGISTLRNSYGAPAEIDIKHNOVSAQVPQEPSPKDYETYPFQFONSTRNY 1026
FFEDQVHHRPPASBYTHMLGLKSTGNVYDFYSAARPYSELNYETSHYPASPDSMW 1225
FFEDQVHHRPPASBYTHMLGLKSTGNVYDFYSAARPYSELNYETSHYPASPDSMW 1084

```

lication US/08982785A

929

ATION:  
Kosik, Kenneth S.  
Zhou, Jianhua

ENTION: ALARM RELATED PEPTIDES AND  
ENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM

QUENCES: 13

CE ADDRESS:

Fish & Richardson P.C.  
25 Franklin Street  
ton

USA

0-2804

DABLE FORM:

E: Diskette

IBM Compatible

SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/982,785A

FILING DATE: 02-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/031,556

FILING DATE: 02-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Freeman, John W.

REGISTRATION NUMBER: 29,066

REFERENCE/DOCKET NUMBER: 05311/018001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 756 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-08-982-785A-2

Query Match 61.4%; Score 3916; DB 3; Length 756;  
Best Local Similarity 94.9%; Pred. No. 4.2e-277;  
Matches 752; Conservative 0; Mismatches 4; Indels 36;

```

QY 434 MRSLSQSGDPLPAHTGTVRTSTAPSSPGVDSVPLQRTSGHQFNAAAATFQF
Db 1 MRSLSQSGVPLPAHTGTVRTSTAPSSPGVDSVPLQRTSGHQFNAAAATFQF
QY 494 GPASNADPYRQYQCPSPSPYSGKSPALPPPGTTLARSPSIDSIQKDPREFGWF
Db 61 GPASNADPYRQYQCPSPSPYSGKSPALPPPGTTLARSPSIDSIQKDPREFGWF
QY 554 EVTOMLQHPSPVSNAAVLOHLCEGDNKIKAEIRROGGIQLLDVLDLHRMTEV
Db 121 EVTOMLQHPSPVSNAAVLOHLCEGDNKIKAEIRROGGIQLLDVLDLHRMTEV
QY 614 GARNLVYKANDDKITALKNCGGIPALVRLRLKTTDLRELVTVGLVNLSSCI
Db 181 GARNLVYKANDDKITALKNCGGIPALVRLRLKTTDLRELVTVGLVNLSSCI
QY 674 IIQDALAVLTNAVIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEE
Db 241 IIQDALAVLTNAVIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEE
QY 734 RECDGLTDALLYVIQALGSSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTI
Db 301 RECDGLTDALLYVIQALGSSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTI
QY 794 LCGEANGKDAESSGCGWKKKKKKSQDQDVGVPDCAEPPKGIQMLWHPISIVK
Db 361 LCGEANGKDAESSGCGWKKKKKKSQDQDVGVPDCAEPPKGIQMLWHPISIVK
QY 854 LSECSNPDTLEGAAGALQNLAAAGSKWKSQVYIRAAVRKEKGLPILVELLRIDNDRV
Db 421 LSECSNPDTLEGAAGALQNLAAAGSKWKSQVYIRAAVRKEKGLPILVELLRIDNDRV
QY 914 TALRNMALDVNRKELIGKYAMRDLVHRLPGGNNNNNTASKAMSDDTVTAVCCTLH
Db 481 TALRNMALDVNRKELIGKYAMRDLVHRLPGGNNNNNTASKAMSDDTVTAVCCTLH
QY 974 NMENAKALRDAGGIEKLVGISKSGDKHSPKVVVKAASQVLSNMWQYRDLRLSLYKX
Db 541 NMENAKALRDAGGIEKLVGISKSGDKHSPKVVVKAASQVLSNMWQYRDLRLSLYKX
QY 1034 YHFAVSSSTIERDRQRPYSSSRTTPSISPRVSPNNRSASAPASPREMISLKERKT
Db 601 YHFAVSSSTIERDRQRPYSSSRTTPSISPRVSPNNRSASAPASPREMISLKERKT
QY 1094 GSNATYHGAKGEHTSRKDAWTAQNTGISTLRNSYGAPAEIDIKHNOVSAQVPQEQE

```

|||||  
TYHGAKGHTSRKDAQNTAGTSTLYR----- 692  
QPFCNSTRNDESEFFQVHRRPPASEYTMHLGKSTGNVYDFYSAARPYSELNVE 1213  
-----NSTRNDESEFFQVHRRPPASEYTMHLGKSTGNVYDFYSAARPYSELNVE 744  
PASPSDW 1225  
PASPSDW 756  
ication US/08982785A  
29  
TION:  
eik, Kenneth S.  
hou, Jianhua  
TION: ALARM RELATED PEPTIDES AND  
NTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM  
UNCES: 13  
E ADDRESS:  
Fish & Richardson P.C.  
5 Franklin Street  
on  
SA  
-2804  
ABLE FORM:  
: Diskette  
IBM Compatible  
YSTEM: Windows 95  
FastSEQ for Windows Version 2.0b  
CATION DATA:  
NUMBER: US/08/982,785A  
: 02-DEC-1997  
TION DATA:  
NUMBER: US 60/031,556  
: 02-DEC-1996  
T INFORMATION:  
nan, John W.  
N NUMBER: 29,066  
CKET NUMBER: 05311/018001  
TION INFORMATION:  
617/542-5070  
17/542-8906  
154  
SEQ ID NO: 8:  
ACTERISTICS:  
6 amino acids  
c acid  
linear  
: protein  
55.8%; Score 3558; DB 3; Length 686;  
arity 99.4%; Pred. No. 4.7e-251;  
onservative 0; Mismatches 4; Indels 0; Gaps 0;  
3DLPAPHTCTYRTSTAPSPGVDSVPLQRTGSHQHPQNAATAATFORASYAAGPAS 497  
GVPLPAPHTCTYRTSTAPSPGVDSVPLQRTGSHQHPQNAATAATFORASYAAGPAS 60  
PYRQLOQCPSPYSKSGPALPEGLTARSPSIDSIQKDPREFGWRDPPELPEVIQ 557  
PYRQLOQCPSPYSKSGPALPEGLTARSPSIDSIQKDPREFGWRDPPELPEVIQ 120  
QFSPVQNAAYLQHLFCGDNKIKAEIRROGGIQLLVLDLDRHMTVEHRSACGALR 617  
QFSPVQNAAYLQHLFCGDNKIKAEIRROGGIQLLVLDLDRHMTVEHRSACGALR 180  
3KANDDNKIALKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALMPLIQD 677

Db 181 NLVYKANDDNKIALKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALMPLIQD 677  
QY 678 ALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLRNVSSAGEARRR  
Db 241 ALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLRNVSSAGEARRR  
QY 738 GLTDALLYVVIQSALGSSSEIDSKTVENCVCILRNLSYRLAAETSOQOHMGTDDELDCI  
Db 301 GLTDALLYVVIQSALGSSSEIDSKTVENCVCILRNLSYRLAAETSOQOHMGTDDELDCI  
QY 798 ANGKDAESSGCGWKKKKKKKSDQDQDVGPLPDCABPPKGIOMLWHPISIVKPYLTI  
Db 361 ANGKDAESSGCGWKKKKKKKSDQDQDVGPLPDCABPPKGIOMLWHPISIVKPYLTI  
QY 858 SNPTTLGAGALONLAAGSMKWSVYIRAAVRKEKGLPIILVELLRINDRNVVCAV  
Db 421 SNPTTLGAGALONLAAGSMKWSVYIRAAVRKEKGLPIILVELLRINDRNVVCAV  
QY 918 NMALDVRNKELIGKYAMRDLVHRLPGGNNNSNTASKAMSDDTVTAVCCTLHEVITP  
Db 481 NMALDVRNKELIGKYAMRDLVHRLPGGNNNSNTASKAMSDDTVTAVCCTLHEVITP  
QY 978 AKALRDAGGIEKLVGISKSGDKHSPKVYVAAASQVLSNMMQYRDLRSLYKKDGWSC  
Db 541 AKALRDAGGIEKLVGISKSGDKHSPKVYVAAASQVLSNMMQYRDLRSLYKKDGWSC  
QY 1038 ASSSTIERDRQRPYSSSRTPTSPISPVVSPNNRSASAPSPREMISSLKERTDYECT  
Db 601 ASSSTIERDRQRPYSSSRTPTSPISPVVSPNNRSASAPSPREMISSLKERTDYECT  
QY 1098 TYHGAKGHTSRKDAQNTAGTSTL 1123  
Db 661 TYHGAKGHTSRKDAQNTAGTSTL 686  
RESULT 4  
US-08-982-785A-10  
; Sequence 10, Application US/08982785A  
; Patent No. 6258929  
; GENERAL INFORMATION:  
; APPLICANT: Kosik, Kenneth S.  
; APPLICANT: Zhou, Jianhua  
; TITLE OF INVENTION: ALARM RELATED PEPTIDES AND  
; TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/982,785A  
; FILING DATE: 02-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/031,556  
; FILING DATE: 02-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Freeman, John W.  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 05311/018001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 10:

## CHARACTERISTICS:

20 amino acids

no acid

linear

E: protein

47.5%; Score 3030; DB 3; Length 620;

Identity 95.6%; Pred. No. 1.4e-212;

Conservative 1; Mismatches 26; Indels 0; Gaps 0;

SOSQDPLPPAHTGYRTSTAPSPGVDSVPLORTGSHGQPNAAAATFORASYAAG 494

SOSQGVLPAPHTGYRTSTAPSPGVDSVPLQRTCSQHCTQNAAAAATFORASYAAG 60

NYADPYRQLOQCPSPVSYSGSPALPEGLTARSIDSIDKDPREFGRWDPELPE 554

NYADPYRQLOQCPSPVSYSGSPALPEGLTARSIDSIDKDPREFGRWDPELPE 120

MLQHPFVSQNAAYLOHLCFGDNKIKAEIRROGGIQLLVLDLDRHMTVHRSAAG 614

MLQHPFVSQNAAYLOHLCFGDNKIKAEIRROGGIQLLVLDLDRHMTVHRSAAG 180

NLVYKANDDNKIALKNCGGIPALVRLKTTDLIELVLTGVLWNLSDDLKMPI 674

NLVYKANDDNKIALKNCGGIPALVRLKTTDLIELVLTGVLWNLSDDLKMPI 240

ALAVLTNAVLIIPHSGWENSPLODDRKIQHSSQVLRNATGCLRVNVSAGBEARRMR 734

ALAVLTNAVLIIPHSGWENSPLODDRKIQHSSQVLRNATGCLRVNVSAGBEARRMR 300

GLTDALLYVQISALGSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTDGLDGL 794

GLTDALLYVQISALGSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTDGLDGL 360

ANGKDAESSGCGWKKKKKSDQDGVPLPDCAEPPKGMWLHPSIVKPYLTLL 854

ANGKDAESSGCGWKKKKKSDQDGVPLPDCAEPPKGMWLHPSIVKPYLTLL 420

SNPDTLECAALQNALAAACEWKEVYIRAAVRKEKGRPIVLVLRINDRVCAVAT 914

SNPDTLECAALQNALAAACEWKEVYIRAAVRKEKGRPIVLVLRINDRVCAVAT 480

NMALDVNRKELIGKYAMRDVLRPLPGGNSNNTASKAMSDDTTAVCCTLHEVITKN 974

NMALDVNRKELIGKYAMRDVLRPLPGGNSNNTASKAMSDDTTAVCCTLHEVITKN 540

AKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLNSMWCYRDLRLSLYKKGWSQY 1034

AKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLNSMWCYRDLRLSLYKKGWSQY 600

ASSTTIERDQRPY 1051

ASSTTIERDQRPY 617

Publication US/09167206A

193

RION:

Jabalan, Krishnan

ulz, Vincent P.

9, Melja

TION: NIK1 PROTEIN AND NIK1 PROTEIN COMPLEXES

: 15966-521 NIK1 protein complexes

ATTION NUMBER: US/09167,206A

DATE: 1998-10-06

ID NOS: 26

ntin Ver. 2.0

; ORGANISM: Homo sapiens

US-09-167-206-14

## Query Match

44.2%; Score 2818.5; DB 4; Length 1211;

Best Local Similarity 49.2%; Pred. No. 1e-196;

Matches 618; Conservative 192; Mismatches 337; Indels 109;

QY 15 MPVPDQPSASASEKTSLSPLNTSNGDGSSETTT-SAILASVKEQELQFRLTRF

DB 1 MPAPQASLVEEQPQTRQEA-ASTGPGMEPETTATILASVKEQELQFRLTRF

QY 74 QIVASOLERCKLGSETGSMSSMSABEQFOWOSQD-----GOKDIEDELTTGLE

DB 60 QIVASOLERCKLGSETGSMSSMSABEQFOWOSQD-----GOKDIEDELTTGLE

QY 128 IRLSLQSSGIL-DFQDYSTGTGPSPLL--SOSALQNSKPEGSFOYPASYHSNQTLA

DB 116 IRTEPQGLIYSPQYSLHSEGLNSRSSTOMNSYSDSGYQEAQSFHNSQNVF

QY 180 -----LGETTPSOLPARGTQARATGQSFQ-GTTSRAGHLAGEPAPPPPP-----

DB 176 QHSHFIGSTNNHV-----RNSRAEGQTLVQPSVANRMRVSSVPSEAQSFSYVI

QY 225 PREPPAPSLGSAFHLP---DAPPAALAAALYYSSSTLPAPPGRGSPPLAAPQGS

DB 232 PSRGLRTSLGSGFGSPVTDPRPLNPSA---YSSTTLPA-ARAAAPY-SQRPAS

QY 282 RGSAPFEGATYAAPRGSSPK-QSPSLAKSYSTSSPINIVVSSAGLSPIRVTSPP

DB 287 RIGSVTSRQT-SNPNGPTQYQTARV-----GSPLTLDAT-----RVASPS

QY 341 ISSSPIHQLSSTIGTYATLSPT-KRLVHASEQYKSHSQELYATATLQRPGLSAAAG

DB 334 GSSSP-KRSGMTAVPQHLGSLQRTVHDMEQPGQQQYDIYERWVPPRDSL-TG

QY 400 SSOHGLGELRALQSPHIDPIYEDRVYQKPPMRSLSSQSQDQDPIPPAHTGYR

DB 391 ASOHSQGLQDLASAVSPDLHTPIYEGRTYYSVYRSPNHGTVE-LQGSOTALYR

QY 460 SSPGVDSV-PLQRTSGHGPQNAATAATFORASYAAGPASNADPYRQLOQCPSPVE

DB 446 ---GVSGIGNLQRTSSQR-----STLTYQRNNLTNTATYAEPIQIRVQ-E

QY 519 SGPALPEGTILARSPSIDSIQKDPREFGRWDPELPEVIQMLQHPFVSQNAAY

DB 497 LQHAVPADDTTSPSIDSIQKDPREFAMKDPPELPEVIHMLEHQFPFVQANAAY

QY 579 FGDNKKABIRQGGITQLLVLDLDRHMTVHRSAOGLRNLYGVKANDDNKIALK

DB 557 FGDNKKABIRQGGITQLLVLDLDRHMTVHRSAOGLRNLYGVKANDDNKIALK

QY 639 PALVRLRKTTLLEIRELVTGVLWNLSDDLKMPIIQDALAVLTNAVLIIPHSGW

DB 617 PALLRLLRKSIDAEVRELVTGVLWNLSDDLKMPIIIRDALSTLNTVIVPHSGW

QY 699 QDRKIQHLSQVLNATCCLRNVSAGBEARRMRRECGLTDALLYVQISALGS

DB 677 DDDHKIKFQTSILVLRNTTGLNLRTSAGEARKQMRSCGELVDSLLYVHTCVNT

QY 759 KTVENCVCILRNLSYRLAETSQGHMGTDGLDGLCGEANGKDAESSGCGWKKK

DB 737 KTVENCVCILRNLSYRLAETSQGHMGTDGLDGLCGEANGKDAESSGCGWKKK

QY 818 --SQDQWQGVLPDCAEPPKGMWLHPSIVKPYLTLLSECSNPDTEGAAGAL

DB 796 TPQEQDQWQGVLPDCAEPPKGMWLHPSIVKPYLTLLSECSNPDTEGAAGAL

QY 876 GSWKSVYIRAAVRKEKGLPIVLVLRINDRVCAVATALRNALDVRNKLIG

DB 856 SNWKEFAVIRGGRPKRKGPIVLVLRINDRVCAVATALRNALDVRNKLIG

QY 936 DLAVHRLPGGNSNNTASKAMSDDTTAVCCTLHEVITKNMENAKALRDAGGIEKL



ton

USA  
0-2804  
DABLE FORM:  
E: Diskette  
IBM Compatible  
SYSTEM: Windows 95  
FastSQ for Windows Version 2.0b

ICATION DATA:

N NUMBER: US/08/982,785A

E: 02-DEC-1997

ATION DATA:

N NUMBER: US 60/031,556

E: 02-DEC-1996

NT INFORMATION:

eman, John W.

ON NUMBER: 29,066

DOCKET NUMBER: 05311/018001

ATION INFORMATION:

617/542-5070

0154

R SEQ ID NO: 9:

RACTERISTICS:

82 amino acids

no acid

linear

E: protein

19.6%; Score 1249.5; DB 3; Length 682;

larity 44.5%; Pred. No. 1.3e-82;

Conservative 105; Mismatches 184; Indels 45; Gaps 11;

SNY-----ADPYRQLOQCPVESFYSGKSGPALPE-----GTLARS 532

SDYGTARRTGTPDERRRLR-----STEDMIGEVEPPDQYWPAPLAQHERGSLA-- 119

DSTQKD--PRFGRDPLPEVIMQLQHPFSVQSNAAAYLQHLCPGDNKIKAEIRRQ 591

DRLKMGPPPSNRKQPELVIAMLGFRDLDAVKSNAAYLQHLCYRNDKVTDAKL 178

QLLVLDLHRMTVEVHRSACGALRNLYVGKANDONKIALKNCGGIPALVRLRLKTTDL 651

PIVLGLLDHPKKEVHLGACGALKNISEFR--DQNKIAIKNCDGVPALVRLRKARDM 237

ELVTGLWNLSSDALKMPIIQDALAVLTNAVLIIPHSGWENSLQDDRKIQHSSQV 711

EVITGTWNLSSHDSIKWEIVDHALHALTDEVIIPHSGWEREPNEDCKPRHIEWESV 297

ATGLRNVSAGAEARRMRRECGLTDALLVYVQSALGSSEIDSKTVENCVCILRNL 771

TAGCLRNVSERSEARKRECDGLVDALFIVQAEIGQKDSKLVENCVCILRNL 357

LAAETSGQHMGTDELDDLCCGANGKDAESSGCGWKK--KKKKSDQDQDVGVGPLP 829

VHREIPQ-----AERYQEAFLPTVANSTGPHASCFGAKKKGKXPTDPANDTVDFP 412

EPKGIQMLHPSITVKPYLLSECSNPDTLEGAAGALQNLAAAGSKWMSYVIRAAVR 889

SPARGVELLQPEVVRVYIISLLKESNTPAILEASAGAIQNLCAGRWYGYIRSAUR 472

GLPILVELLRIDNDVVCAVATARNMALDVNRNKLIGKYMARDLVRHLPGGNNSN 949

ALSARAEELLTSQHEVRVVAASGALRNLAVDARNKELIGKGAIRPNLVKNLPGGQON-- 530

KAMSDDDTWTAVCTTLEHVIITKNMENAKALDADGIEKLVGISKSGKHSPKVVKAA 1009

WNSEDIVVILNTINEVIAENLAAKLRBTQIEKLVINKS--GNRSEKEVRAA 587

LNSMWQYRLRSLYKDGQWSQYHFVASSSTIERDR--QRPYSSRTFSPISPRVSPNN 1068

Db 588 ALVLQITMGYKELRKPLEKEGWKSDFOVNLNNAERSQSSHSYDDSTLPLIDRNC

QY 1069 RS 1070

Db 648 YS 649

## RESULT 8

US-08-982-785A-11

; Sequence 11, Application US/08982785A

; Patent No. 6258929

; GENERAL INFORMATION:

; APPLICANT: Kosik, Kenneth S.

; APPLICANT: Zhou, Jianhua

; TITLE OF INVENTION: ALARM RELATED PEPTIDES AND

; TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows 95

; SOFTWARE: FastSEQ for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/982,785A

; FILING DATE: 02-DEC-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/031,556

; FILING DATE: 02-DEC-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Freeman, John W.

; REGISTRATION NUMBER: 29,066

; REFERENCE/DOCKET NUMBER: 05311/018001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 666 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-982-785A-11

Query Match

11.0%; Score 702.5; DB 3; Length 666;

Best Local Similarity 31.3%; Pred. No. 1e-42;

Matches 201; Conservative 97; Mismatches 253; Indels 91;

QY 447 PAHTGYRTSTAPSPGVDSVPLQRTGSOHQGNAAAAT-----FQASVAA

Db 43 PIYNGTLKRE--PDNRRFSSYSQMNWRHYPGSCNITGAGSDICFMQIKASR

QY 499 YADP-----YRQLOQCPVESPYSS-----KSGPALPEGTLA

Db 101 YCDPRGTLRKGLSGKQKQTKQWRY-----SFYSTCSGQKAIKKCPVRPFCASK

QY 536 ----DSIQKDPREFGW-----RDPE-----LPEVIQMLQHPFSVQSNAAAY.

Db 156 IPPISCNKD--LSFGWSRASSKICSEDIESCILTIKAVQYLEQDEKYOAGIAYY

QY 579 FGDNKKIKAEIRQGGIQLLVLDLHRMTVEVHRSACGALRNLYVGKANDONKIALK

Db 215 FODESAKQVYQGGICKLVLDLRSFNQVQQAACALRNLFVRETT--NKLETR

QY 639 PALVRLRLKTTDLREILVTGVLWNLSSCDALKMPIIQDALAVLTNAVLIIPHSGW





3SKLAKKRA D - - - LVEKVRREAEQ RARE 807

us-09-501-171a-4.rapb

Sequence 1  
Sequence 2  
Sequence 3  
Sequence 4  
Sequence 5  
Sequence 6  
Sequence 7  
Sequence 8  
Sequence 9  
Sequence 10  
Sequence 11  
Sequence 12  
Sequence 13  
Sequence 14  
Sequence 15  
Sequence 16  
Sequence 17  
Sequence 18  
Sequence 19  
Sequence 20  
Sequence 21  
Sequence 22  
Sequence 23  
Sequence 24  
Sequence 25  
Sequence 26  
Sequence 27  
Sequence 28  
Sequence 29  
Sequence 30  
Sequence 31  
Sequence 32  
Sequence 33  
Sequence 34  
Sequence 35  
Sequence 36  
Sequence 37  
Sequence 38  
Sequence 39  
Sequence 40  
Sequence 41  
Sequence 42  
Sequence 43  
Sequence 44  
Sequence 45  
Sequence 46  
Sequence 47  
Sequence 48  
Sequence 49  
Sequence 50  
Sequence 51  
Sequence 52  
Sequence 53  
Sequence 54  
Sequence 55  
Sequence 56  
Sequence 57  
Sequence 58  
Sequence 59  
Sequence 60  
Sequence 61  
Sequence 62  
Sequence 63  
Sequence 64  
Sequence 65  
Sequence 66  
Sequence 67  
Sequence 68  
Sequence 69  
Sequence 70  
Sequence 71  
Sequence 72  
Sequence 73  
Sequence 74  
Sequence 75  
Sequence 76  
Sequence 77  
Sequence 78  
Sequence 79  
Sequence 80  
Sequence 81  
Sequence 82  
Sequence 83  
Sequence 84  
Sequence 85  
Sequence 86  
Sequence 87  
Sequence 88  
Sequence 89  
Sequence 90  
Sequence 91  
Sequence 92  
Sequence 93  
Sequence 94  
Sequence 95  
Sequence 96  
Sequence 97  
Sequence 98  
Sequence 99  
Sequence 100

## ALIGNMENTS

```

RESULT 1
US-10-071-900-3
; Sequence 3, Application US/10071900
; Publication NO. US20020127541A1
; GENERAL INFORMATION:
; APPLICANT: St. George-Hyslop, Peter H.
; APPLICANT: Rommens, Johanna
; APPLICANT: Fraser, Paul E.
; TITLE OF INVENTION: Alzheimer's Related Proteins and Methods
; TITLE OF INVENTION: Of Use
; FILE REFERENCE: 1034/1P810-US1
; CURRENT APPLICATION NUMBER: US/10/071,900
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US/09/227,725
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Fastseq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (115)
; OTHER INFORMATION: Xaa can be any amino acid
; US-10-071-900-3

```

Query Match	88.2%	Score 5625	DB 13	Length 1084
Best Local Similarity	99.7%	Pred. No. 0		
Matches 1075	Conservative	0	Mismatches 3	Indels 0
G				
QY	148	PSLLSQSALQLNKPEGSFOYPASHYHSNQTALGTTTSPQLPARGTQARATGQSFS		
Db	7	PSLLSQSALQLNKPEGSFOYPASHYHSNQTALGTTTSPQLPARGTQARATGQSFS		
QY	208	SRAGHLGAPGPAPPPPPPPPPAPSLGSAFHLPDAPPAAMAAALYYSSSTLPAPP		
Db	67	SRAGHLGAPGPAPPPPPPPPPAPSLGSAFHLPDAPPAAMAAALYYSSSTLPAPP		

the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, based on analysis of the total score distribution.

## SUMMARIES

ery	chn	Length	DB	ID	Description
8.2	1084	13	US-10-071-900-3	Sequence 3, Appl	
4.2	1211	14	US-10-242-943-14	Sequence 14, Appl	
1.8	1110	14	US-10-149-819-12	Sequence 12, Appl	
1.4	1193	13	US-10-071-900-4	Sequence 4, Appl	
6.0	785	15	US-10-264-049-2968	Sequence 2968, Ap	
1.4	941	14	US-10-205-823-80	Sequence 80, Appl	
0.1	834	15	US-10-094-749-2227	Sequence 2227, Ap	
0.7	425	9	US-09-925-301-1274	Sequence 1274, Ap	
7.4	86	9	US-09-864-761-39492	Sequence 39492, A	
7.1	198	15	US-10-094-749-1831	Sequence 1831, Ap	
6.2	164	14	US-10-029-348-33760	Sequence 33760, A	
4.4	55	9	US-09-864-761-40085	Sequence 40085, A	
3.9	50	9	US-09-864-761-38977	Sequence 38977, A	
3.7	44	9	US-09-864-761-39938	Sequence 39938, A	
3.6	1832	14	US-10-128-714-8298	Sequence 8298, Ap	

APOGGSPTKLQSGSAPEGATYAAPRGSSPKQSPRLAKSYSTSSPINIVVSSAGLS 327  
 |||||  
 APOGGSPTKLQSGSAPEGATYAAPRGSSPKQSPRLAKSYSTSSPINIVVSSAGLS 186  
 |||||  
 VTPPTVQSTISSPIHQLSSTIGTYATLFTKELVHASQYKSHSOELVATATLQR 387  
 |||||  
 VTPPTVQSTISSPIHQLSSTIGTYATLFTKELVHASQYKSHSOELVATATLQR 246  
 |||||  
 LAAGSRASYSGHGLPELALQSPHEHIDPIYEDRVYQKPPMRSLSQSGDPLPP 447  
 |||||  
 LAAGSRASYSGHGLPELALQSPHEHIDPIYEDRVYQKPPMRSLSQSGDPLPP 306  
 |||||  
 3TTRTAPSSPGVDSVFLQRTSQHQGPQNAATAATFORASYAAGPASNYADPYRQLQ 507  
 |||||  
 3TTRTAPSSPGVDSVFLQRTSQHQGPQNAATAATFORASYAAGPASNYADPYRQLQ 366  
 |||||  
 SVESPYSKSPALPPEGTTLARSPIDSIOKDPREFGHRDPELVEIOMLOHQPFVSQ 567  
 |||||  
 SVESPYSKSPALPPEGTTLARSPIDSIOKDPREFGHRDPELVEIOMLOHQPFVSQ 426  
 |||||  
 AAYLQHLCFGDNKIKABIRQGGIQLLVLDLDRHMTVEVHRSACGALRNLYGKANDD 627  
 |||||  
 AAYLQHLCFGDNKIKABIRQGGIQLLVLDLDRHMTVEVHRSACGALRNLYGKANDD 486  
 |||||  
 ALKNCGGIPALVRLRLKTTDLTRELVTGVLWNLSLSCDALKMPIIQDALAVLTNAV 687  
 |||||  
 ALKNCGGIPALVRLRLKTTDLTRELVTGVLWNLSLSCDALKMPIIQDALAVLTNAV 546  
 |||||  
 SGWENSPLODRKIQHSSQVLRNATGCLRNVSAGAEARRMRECDGLTDALLYVI 747  
 |||||  
 SGWENSPLODRKIQHSSQVLRNATGCLRNVSAGAEARRMRECDGLTDALLYVI 606  
 |||||  
 JGSEIDSKTVENCVCILRNLSYRLAAETSGQGHMGTDGLLGCENGKDAESSG 807  
 |||||  
 JGSEIDSKTVENCVCILRNLSYRLAAETSGQGHMGTDGLLGCENGKDAESSG 666  
 |||||  
 KKKKKKSQDQWGVGLPDCAPPKGIQMLWHPISIVKPYLTLLSECSNPDITLEGAA 867  
 |||||  
 KKKKKKSQDQWGVGLPDCAPPKGIQMLWHPISIVKPYLTLLSECSNPDITLEGAA 726  
 |||||  
 QNLAAGSKWKSIVIRAAVRKEKGLPILVELLRIDNDRVVCVATALRNALDVRNKE 927  
 |||||  
 QNLAAGSKWKSIVIRAAVRKEKGLPILVELLRIDNDRVVCVATALRNALDVRNKE 786  
 |||||  
 KYAMRDLVHRLPGGNNNSNTASKMSDDTVAVCCTLHEVITKNMENAKALRDAGGI 987  
 |||||  
 KYAMRDLVHRLPGGNNNSNTASKMSDDTVAVCCTLHEVITKNMENAKALRDAGGI 846  
 |||||  
 JGTSKSGDKGSKPKVYKAAQVLSNMWQYDRLSLYKKGWSQYHFAVSSSTIERDR 1047  
 |||||  
 JGTSKSGDKGSKPKVYKAAQVLSNMWQYDRLSLYKKGWSQYHFAVSSSTIERDR 906  
 |||||  
 JSSSRTPSIPVRVSPNNRSASAPASPREMISLKERKTDYECTGNSNATYHGAKGEHT 1107  
 |||||  
 JSSSRTPSIPVRVSPNNRSASAPASPREMISLKERKTDYECTGNSNATYHGAKGEHT 966  
 |||||  
 JAMTAQNTGSTLYRNSYGAPAEIDIKHNQVSAQVPQEPKRDYETYPQPNSTRNY 1167  
 |||||  
 JAMTAQNTGSTLYRNSYGAPAEIDIKHNQVSAQVPQEPKRDYETYPQPNSTRNY 1026  
 |||||  
 FFEQDVHRRPPASEYTMHLGKSTGNVVDYFSAARPYSELNYETSHYPASPDWSW 1225  
 |||||  
 FFEQDVHRRPPASEYTMHLGKSTGNVVDYFSAARPYSELNYETSHYPASPDWSW 1084  
 |||||

ollication US/10242943  
 US2003008/412A1  
 IION:  
 Iabalan, Krishnan  
 uliz, Vincent P.

; APPLICANT: Yang, Meija  
 ; TITLE OF INVENTION: NIK1 PROTEIN AND NIK1 PROTEIN COMPLEXES  
 ; FILE REFERENCE: 15966-521 NIK1 protein complexes  
 ; CURRENT APPLICATION NUMBER: US/10/242,943  
 ; CURRENT FILING DATE: 2002-09-13  
 ; PRIOR APPLICATION NUMBER: US/09/167,206  
 ; PRIOR FILING DATE: 1998-10-06  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 14  
 ; LENGTH: 1211  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-242-943-14

Query Match 44.2%; Score 2818.5; DB 14; Length 1211;  
 Best Local Similarity 49.2%; Pred. No. 1.4e-177;  
 Matches 618; Conservative 192; Mismatches 337; Indels 109;

QY 15 MPVPDQPSASBKTSLSPLNTSNGDSETETT-SAILASVKEQELQFORLTRE  
 |||||  
 Db 1 MPAEQASLVERGQPTRQEA-ASTGPGMEPETTATTILASVKEQELQFORLTRE  
 |||||  
 QY 74 QIVASQLERCKLGETGSMSSMSAEQFQWOSQD-----GKQIEDLBTGLE  
 |||||  
 Db 60 QIVASQLERCKLGAESPSTASTSTEKSPMRSTDPVNTGVSKPRVSDAVQPNY  
 |||||  
 QY 128 ISLSQESGTL-DPODYSTGTERPSLL--SOSALQNSKPEGSFOYPASVHSNOTLA  
 |||||  
 Db 116 IITEFQGLTYFEQTSLSHESEGLNSRSSTOMNSYSDSGYQEAGSPNSQVNS  
 |||||  
 QY 180 -----LGETTPSQLPARGTQARATQGSFSQ-GTTSRAGHLAGEPAPPPPP  
 |||||  
 Db 176 QQHSFTGSTNNEHW---RNSRAEGQTLVQPSVANRRRVSSVPSRAQSPSVI  
 |||||  
 QY 225 PREPEFAPSLGSAFHLP---DAPPAALAAALYSSSTLPAPPGSGPLAAPQGS  
 |||||  
 Db 232 PRGSLRSLGSGFGSPSVTDPRPLNPSA---YSSSTLPA-ARAAAPY-SQRPAS  
 |||||  
 QY 282 RGSAPGATYAAPRGSSPK-QSPRLAKSYSTSSPINIVSSAGLSPIRVTSPP  
 |||||  
 Db 287 RIGSVTSRGT-SNPNPTFOYQTTARV-----GSPLTLDATQ-----RVASPS  
 |||||  
 QY 341 ISSSPIHQLSSTIGTYATLSP-T-KRLVHASQYKSHSOELVATATLQRPGLAAG  
 |||||  
 Db 334 GSSSP--KRSGMTAVPQHLGSLQRTVHDMEQGQQQYDIYERVMVPRPDSL-TG  
 |||||  
 QY 400 SSOHGLGPELALQSPHEHIDPIYEDRVYQKPPMRSLSQSGDPLPPAHTCTYR  
 |||||  
 Db 391 ASQHSLGQDLKSAVSPDLHTPIYEGRTYSPVYRSPNHTVE-LQGSQTALYR  
 |||||  
 QY 460 SSFGYDSV-PLQRTSQHQGPQNAATAATFORASYAAGPASNYADPYRQLQCPVSVE  
 |||||  
 Db 446 --GVSGIGNLQRTSQR-----STLTQYNNYALNTTATYAEYRPIYRQ-E  
 |||||  
 QY 519 SGPALPPEGTTLARSPIDSIOKDPREFGHRDPELVEIOMLOHQPFVSQNSNAAA  
 |||||  
 Db 497 LQHAVPADDTTRSPIDSIOKDPREFAMRDPELVEIOMLOHQPFVSQNAAAA  
 |||||  
 QY 579 FGDNKKABIRQGGIQLLVLDLDRHMTVEVHRSACGALRNLYGKANDDNKIALKI  
 |||||  
 Db 557 FGDNKKVKEVCLGGIKHLVLDLDRHVEVQKNACGALRNLYGKSTDENKIAMKI  
 |||||  
 QY 639 PALVRLRLKTTDLTRELVTGVLWNLSLSCDALKMPIIQDALAVLTNAVIIHSGWI  
 |||||  
 Db 617 PALVRLRLKSIDAEVRELVTGVLWNLSLSCDAVKMTIIRDALSTLTNTVIVPHSGMI  
 |||||  
 QY 699 QDRKIQHSSQVLRNATGCLRNVSAGAEARRMRECDGLTDALLYVIQSLGSG:  
 |||||  
 Db 677 DDDHKIKTQTSVLRLNTTGCLRNLTSGAEARKQMRSCGELVDSLTYVHTCVNT:  
 |||||  
 QY 759 KTVENCVCILRNLSYRLAAETSGQGHMGTDGLLGCENGKDAESSGCKWKKKI  
 |||||



```

; NAME/KEY: MISC_FEATURE
: LOCATION: (769)

```

```

RESULT 5
US-10-264-049-2968
/ Sequence 2968, Application US/10264049
/ Publication No. US2004000579A1
/ GENERAL INFORMATION:
/ APPLICANT: Birse et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PA133P1
/ CURRENT APPLICATION NUMBER: US/10/264,049
/ CURRENT FILING DATE: 2002-10-04
/ PRIOR APPLICATION NUMBER: PCT/US01/18569
/ PRIOR FILING DATE: 2001-06-07
/ PRIOR APPLICATION NUMBER: US 60/209,467
/ PRIOR FILING DATE: 2000-06-07
/ NUMBER OF SEQ ID NOS: 4360
/ SOFTWARE: PatentIn Ver. 3.1
/ SEQ ID NO 2968
/ LENGTH: 785
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (482)
/ OTHER INFORMATION: Xaa equals any of the twenty naturally occurri
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (636)
/ OTHER INFORMATION: Xaa equals any of the twenty naturally occurri
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (768)

```

14:26:19 2004

us-09-501-171a-4.rapb

ION: Xaa equals any of the twenty naturally occurring L-amino acids

FEATURE

ION: Xaa equals any of the twenty naturally occurring L-amino acids

FEATURE

ION: Xaa equals any of the twenty naturally occurring L-amino acids

FEATURE

ION: Xaa equals any of the twenty naturally occurring L-amino acids

36.0%; Score 2296; DB 15; Length 785;  
arity 58.8%; Pred. No. 3e-143;  
nservative 117; Mismatches 173; Indels 34; Gaps 14;

PTVQSTISSPIHQLSSITGTATLSPT-KRLVHASEQYKSHSQELYATATLQRP 388

PSOQ-VGSSP--KSGMTAVPQHLGSLQRTVHDMEQFGQQQYDIYERMPRP 64

AGRAYSQHGHGPELRALQSPHHIDPIYEDRVYKPPMRSLQSQGDPLPPA 448

GLRSSVYASQHSQGLQDLRSVSPDLHITPIYEGRTYSPVYRSPNHGTVE-LQGS 122

RTSTAPSPGVDSVPLORTGSHGQCNAAATFORASYAAGPASYADPYRLOY 508

RT-----GSGVIGN--LQRTSSOR-----STLYQRNNYALNTATYAEYRPIQY 171

SPYSKGPALPEGTIARSPSIDSIQKOPREFGWRDPPELPIVQIOMLQHOFFSVQS 568

ICNYNLQHAVPADGTTSPSIDSIQKOPREFAMRDPPELPIVHMLQHOFFSVQA 230

LQHLCEGDKNIABTRQGGIQLVLDLDRHMTVEHRSACGALRNLYVCKANDNN 628

LQHLCEGDKNVMKVCRLGKHLVDLDRHMLEVQKNACGALRNLYVCKSTDEN 290

QNGGGTALVRLAKRTDLEIRELVTGLVNLSSCDALKMPITQDALAVLTNAVII 688

QNGGGTALVRLAKRTDLEIRELVTGLVNLSSCDALVNTNAVII 350

VENSPLQDDRKIQHSSQVLNRNATGCLRNVSAGAEARRRMRCDGLTALLYVIQ 748

VNNSPDDDKIKFQISLVLRNTTGLCLRNLSAGAEARKQMRCEGLVDSLLVVIH 410

3SIDSKTVENCVCILRNLSYRLAAETSCQOHMGTDLDGLLCEANGKDAESSGC 808

TSYDSKTVENCVCILRNLSYRLAAETSCQOHMGTDLDGLLCEANGKDAESSGC 469

QKKK--SODQDVGVPDCAEPKGIOMLWHPSTVKPYLTLLSECSNPDFTLEG 865

QKKKTPXEDQDVGVPDCAEPKGIOMLWHPSTVKPYLTLLSECSNPDFTLEG 529

QNLASGKMWKSVYIRAAVKEKGLPLTVELLRIDNDRVVCAVATALRNALDVRN 925

QNLASGKMWKSVYIRAAVKEKGLPLTVELLRIDNDRVVCAVATALRNALDVRN 589

3KAMRDLVRLPGGNNNTASKAMDDTAVTCCTLHVITKNNENAKALRDAG 985

3KAMRDLVRLPGGNNNTASKAMDDTAVTCCTLHVITKNNENAKALRDAG 644

LVGISKGDKHSPKVYKAAQSVLNSMWOYRDLRSYKKGWQYHFVASSSTIER 1045

LUNITKGRDRSSLKVYKAAQSVLNSMWOYRDLRSYKKGWQYHFVASSSTIER 704

PYSSRTPTSPV--RVSPNNRSASAPSPREMIILKERTDYECTGSNATYHGAK 1103

SH-----PSLSTTNQOMSPIIQSVGSTSSPALLGIRDPREYDRTPQPMQYNSQ 759

SR 1109

Db 760 GDATHK 765

RESULT 6

US-10-205-823-80

; Sequence 80, Application US/10205823

; Publication No. US20030108963A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Monahan, John E.

; APPLICANT: Endege, Wilson O.

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Gorbacheva, Bella

; APPLICANT: Hoersch, Sebastian

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Wensey, Angela M.

; APPLICANT: Glatt, Karen

; APPLICANT: Zhao, Xumei

; APPLICANT: Anderson, Dustin

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND

; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVEN

; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER

; FILE REFERENCE: MRI-044

; CURRENT APPLICATION NUMBER: US/10/205,823

; CURRENT FILING DATE: 2002-07-25

; PRIOR APPLICATION NUMBER: 60/307,982

; PRIOR FILING DATE: 2001-07-25

; PRIOR APPLICATION NUMBER: 60/314,356

; PRIOR FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/325,020

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: 60/341,746

; PRIOR FILING DATE: 2001-12-12

; PRIOR APPLICATION NUMBER: 60/362,158

; PRIOR FILING DATE: 2002-03-05

; NUMBER OF SEQ ID NOS: 455

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 80

; LENGTH: 941

; TYPE: PR

; ORGANISM: Homo sapiens

; US-10-205-823-80

Query Match 21.4%; Score 1362; DB 14; Length 941;

Best Local Similarity 33.1%; Pred. No. 3e-81;

Matches 357; Conservative 162; Mismatches 323; Indels 236; G

Qy 41 DGSSETTSAILASVKEQELQFRLTRELBAERQIVASQLERCKLG-SETGSMSSM

Db 2 DDESEVESTASILASVKEQEAQFEKLTALREBERHVSQAQLERVRVSPQDANPLMAN

Qy 100 EQFQWQSQDQKDIETLTGLVDSCIRSLQESGILDPQDYSTGERPSILLSQSA

Db 62 RHQNGRFVGDADLEKQKPSDLKL-----NG---QDHS-----LYST

Qy 160 SKPEGSFQYPASYHSNQTALGETTFSQLPARGTQAPATQCSFSGTTSRAGHLAG

Db 103 QEPG-----QIVETYTEEDP-EGAMSVSVSVETSDDGTTTTRT-----

Qy 220 PPPPPPPPPFPAPSLGSAFHLDPAPAAAAAALYYSSSTLPAPPGRGSPALAPCGG

Db 138 -----ETIVKVKVTVTTRTVQPVAMGPDGL-----P

Qy 280 LQSGSAGEATYAAPRGSSPKQSPSLAKSYSTSPINIWSSAGLSPIRVTSP

Db 138 -----IGTVATLSPTKRLVHASEQYKSHSQELYATA

Qy 340 TISSSPHQLSS-----IGTVATLSPTKRLVHASEQYKSHSQELYATA

Db 169 SVNNYIQTILGRDFRKNNGGPGYVGVQAGTATLPRNFHYPPDGYSKHYEDGY---

Qy 388 PGSLAAGRASYSQGHGLGP-----ELRALQSPHHIDPIYEDRVYKQKPMRSLQ

-----SDNYGSLRVTRIEBRYRPSMEGYRAPSQD-VYGPQPVVRVGGSSVD 271  
PAHGTGTRTSTAPSPGVDSV-----PLQRTGSHQHPQNAATAATFORASYAAG 494  
RFHPEPYGLEDDQDSMGVDDLDYGMMSDYGTRARTGT----- 311  
NYADPYQLOVCPVESPYSGP-----ALPPEGTIARSPSIDSTOK-DPRE 544  
--SDPRRLR---SYEDMIGEVSDQYVWAPLAQHERGSLA---SLDSURKGGPPP 360  
RDPELPEVIQMLQHPFVSQNAAYLQHLGFGDNKIKAEIRROGGIQLVLDLDRH 604  
RQPELPEVIAMIGFELDAVKSNAAYLQHLGFCYRNDKVKTVRKLKGPVLVGLLDHP 420  
VHRSACALRNLYVKGANDNDKIALKNCGGIPALVRLRLKTTDLIEIRLVTGVLANL 664  
VHLGACGALKNISGR-DQDNKIAIKNCDGVPALVRLRLKARDMDLDEVITGTLANL 479  
DALKMP1I0DALAVLTNAVIIPHSWGNSPLQDDRK1QLHSSQVLRNATCLRNVS 724  
DSIKMEIVDRHALHALTDEVIIIPHSGWREPNEDCKPRHIEWESVLTWTAGCLRNVS 539  
EARRMRECDGLTDALLYVQSALGSSEIIDSKTVCNVCILRNLSYRLAAETSQGH 784  
EARKLRECDGLVDALFIVQAEIGQKSDSKLVENCVCILLRNLSYQVHREIPQ--- 596  
DELGLLCEANGKADSSGCGWKKKKKKK--QDQWDGVGPLPDCAEPKQIMLW 842  
ERYQEAAPNVANNTGPHAAFCGAKGKPKPIEDPANDTVDFPKRTSPARGVELLF 654  
IVKPYLLGECSPDPTLEGAAGALQNLAAAGSKWWSVYIRAAVKEKGLPLVELLR 902  
VVRIYISLLESKTPALIEASAGAIQNLGAGRTYGYIRSAURQEKALSAIADLLT 714  
DRVCAVATALRNMALDVNRKELIGKYAMRDLVHRLFCGNNNSNNTASKAMSDDTVTA 962  
ERVVKAASGALRNLAVDARNKELIGKHAIPNLVKNLPGQCN---SSWNFSEDTIVS 771  
TLHEVITYNNENAKALDAGGIEKLVGISKSGDKHSPKVVKAAASQVLSNMWQYRDL 1022  
TINEVIAENLFAAKKURETOGIEKLVLINKS--GNRSEKEVRAAALVLOTWINGYKEL 829  
YKDGWQYHFWASSSTIETR-ORPYSSSRTPTSPVVRVSPNNRSASAPASPRE 1079  
LEKGGKKSDFQVNLNNAASQSSHSYDDSTLPLIDRNQKSDNYS-----TENE 882

Application US/10094749  
US20030219741A1

TIION: TAKAO

GIYAMA, TOMOYASU

SUKI, TETSUJI

KAMATSU, AI

TO, HIROYUKI

HII, SHIZUKO

MAMOTO, JUN-ICHI

ONO, YUUKO

O, YURI

SUKA, KAORU

GAI, KEIICHI

IE, RYOTARO

MECHIKA, ICHIRO

KI, NAOHICO

SHIKAWA, TSUTOMU

SUKA, MOTYUKI

GAHARI, KENJI

SUHO, YASUHIKO

TIION: NOVEL FULL-LENGTH cDNA

: 084335/0160

ATION NUMBER: US/10/094,749

CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2227  
LENGTH: 834  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-094-749-2227

Query Match 20.1%; Score 1282.5; DB 15; Length 834;  
Best Local Similarity 45.1%; Pred. No. 4,7e-76;  
Matches 321; Conservative 94; Mismatches 220; Indels 77;

QY 13 GAMPVDPQSSASEKTSLSPLNTSGDSETETT-SAILASVKBOELQFBLT  
DB 130 GMPAPEQASLVEEGOPOTRQEA-ASTGPMETTTATILASVKEQELQFBLT  
QY 72 ERQIVASQLERCKLSETGSMSSMAERQFQWQSQD-----GQKDI EDELTTG  
DB 189 ERQIVASQLERCKLGAESPASIASTSTSKSPWRSTDVENTGSKPRVSDAVQPN  
QY 126 SCIRSLQESGIL-DPQDYSTGERPSSLQSALQNLKSKPEGSFOYPASYHSNQTLA  
DB 247 --IRTEPEQGTLYSPQTSLHERSLGNSRSSTOMNSYSDSGYQEAQSFHNSQNV  
QY 180 ----LGETTPTSPQARGTQARATGQSFQ-GTTSRAGHLAGEPAPAPPPP-----  
DB 305 QHSHFSGTNNHV----RNSRAEGQTLVQPSVANRPMRVSSVPSRAQSPSVI  
QY 225 PREPPAPSLGSAFHLP---DAPPAALAAALYYSSSTLPAPPGRGSLAQAQGS  
DB 361 PGRSLRTSLGSGFGSPVTDPRPLNPSA---YSSSTLPA-ARAASPY-SQRPAS  
QY 282 RGSAGEGATYAAPRGSSPK-QSPSLAKSYSTSSPINIVVSSAGLSPIRVTSPE  
DB 416 RIGSVTSROT-SNPNGPTFOYTARV-----GSPLTITDAQT----RVASPS  
QY 341 ISSPSIHLQSSITGTATLSPT-KRLVHASEQYKSHSQELYATATILQRPGLAAG  
DB 463 GSSSP--KRSGMTAVPQHLGSLQRTVHDMEQFGQQQYDIYERWVPRDLSL-TG  
QY 400 SSOHGLGPELRALQSPHHIDPIYEDRVYQKPPMRLSOSQGDPLPPAHTGYR  
DB 520 ASQHSQLGQDLRSVSPDLHTPIYEGRTYSPVYRSPNHGTVE-LQGSQTALYR  
QY 460 SSPGVDSVPLQRTGSHQHPQNAATAATFORASYAAGPASVADPYRQLOVCPVES  
DB 575 GSVIGN--LQRTSSQ-----STLTQRNNYALNTATYAEYRPIQTRVQ-EC  
QY 520 GPALPPEGTIARSPSIDSIQKDPREFGWRDPDELPEVIOMLQHPFVSQNAAYL  
DB 627 QHAVPADDCGTTSPSIDSIQKDPREFAWRDPDELPEVIHMLQHPFVSQNAAYL  
QY 580 GDNKIKAETRRGGIQLIIVDLDRHMTVEVRSACGALRNLYVKGANDNDKIALKN  
DB 687 GDNKVMEVCRLLGGIKHLVDLDRHVLVEQKNAAGALRNLYVKGSTDENKIAMKN  
QY 640 ALVRLRKTTDLIEIRLVTGVNLNLSGCCDALKMP1I0DALAVLTNAVIIPHS 69  
DB 747 ALLRLKRSIDAEVRELVTGVNLNLSGCCDAVKGTTIIRDALST-THSVLWKHN 79

## RESULT 8

US-09-925-301-1274

; Sequence 1274, Application US/09925301

; Patent No. US20020052308A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies



PA106  
 TION NUMBER: US/09/925,301  
 DATE: 2001-08-10  
 ON NUMBER: PCT/US00/05882  
 TE: 2000-03-08  
 ON NUMBER: 60/124,270  
 TE: 1999-03-12  
 D NOS: 1694  
 tIn Ver. 2.0

sapiens

ION: Xaa equals any of the naturally occurring L-amino acids

arity 10.7%; Score 684; DB 9; Length 425;  
 arity 38.8%; Pred. No. 8e-37;  
 conservative 76; Mismatches 144; Indels 20; Gaps 7;

EEARRRRECDGLDALLYVLOSALGSEIDSKTVENCVCILRLNLSRLAAETSGQ 782  
 SEARRKRECDGLDALIFIVQABIGQKDSKLVENCVCILRLNLSYQVHREIPQ- 59  
 TDELGLLGGANGKDAESSGCKKKKKKKS--QDQDVGVPDPCAEPPKGIOM 840  
 AERYQEAAPNVANNTPHAAACFAKKGKPPEDPANDTVDFPKTSPARGVEL 115  
 SIVKPYTLSECSNPDTLEGAAGALONLAGSKWSYVYIRAAVRKEKGLPIIVEL 900  
 EVVRIYISLLKESKTPAILLEASAGAIQNLACGRWYGYKYSALRQKALSIAIDL 175  
 NDRVVCVATARNMALDVNKEKELIGKVMRDVHRLPGGNNNSNTASKMSDDTV 960  
 HERVVKAASGALRNLDVARNKELIGHAIPNLVKNLPGGQON---SSWNPSEDTV 232  
 CTLHEVITKXENAKALRDAGGIEKLVGSKSGDKHSPKVVKAAASOVLSNMQYR 1020  
 NTINEVTAENLEAAKKLRETCQIEKVLINKS--GNRSEKEVRANALVLQTIWYK 290  
 LYKDGQSHVFASSSTIERDR-QRPYSSRTSPISFVRVSPNMRASAPSPRE 1079  
 PLEKEGKKSDFOVNLNNAKSSQSHSYDDSTLPLI-----DRNQKSDKEP--DRE 343  
 KERKTDVECTGSHATYHGAKEHTSRKD 1111  
 SNMGSTKSLDNNVSTPNERGDHNRITLD 375

Application US/09864761  
 20048763A1

ION:

Sharon G.

David R.

el, David K.

Wensheng

ION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ION: GENE EXPRESSION ANALYSIS BY MICROARRAY

Acemica-X-1

ION NUMBER: US/09/864,761

DATE: 2001-05-23

ON NUMBER: US 60/180,312

TE: 2000-02-04

ON NUMBER: US 60/207,456

TE: 2000-05-26

ON NUMBER: US 09/632,366

TE: 2000-08-03

ON NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 39492  
 ; LENGTH: 86  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC005610.1  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.3  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.5  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2  
 ; OTHER INFORMATION: SWISSPROT HIT: Q10145, EVALUATION 1.60e+00  
 US-09-864-761-39492

Query Match 7.4%; Score 475; DB 9; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-24;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; G

QY 1140 VSAQPVQEPSRKDYETYPQFNSTRNYDESFEDQVHHRPPASEYTMHLGLKSTG  
 Db 1 VSAQPVQEPSRKDYETYPQFNSTRNYDESFEDQVHHRPPASEYTMHLGLKSTG  
 QY 1200 FYSAARPYSELNYETSHYPASPDWV 1225  
 Db 61 FYSAARPYSELNYETSHYPASPDWV 86

RESULT 10  
 US-10-094-749-1831  
 ; Sequence 1831, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI

RESULT 12  
US-09-864-761-40085  
: Sequence 40085, Application US/09864761  
: Patent No. US20020048763A1  
: GENERAL INFORMATION  
: APPLICANT: Penn, Sharon G.  
: APPLICANT: Rank, David R.  
: APPLICANT: Rank, David R.  
: APPLICANT: Rank, David R.

```

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;
; FILE REFERENCE: Aeomica-X-1
;
; CURRENT APPLICATION NUMBER: US/09/864,761
;
; CURRENT FILING DATE: 2001-05-23
;

```

```

/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/006666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006664

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668

```

```

/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21

```

```

; PRIOR APPLICATION NUMBER: US 09/774,203
;
; PRIOR FILING DATE: 2001-01-29
;
; NUMBER OF SEQ ID NOS: 49117
;
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;
; SEQ ID NO 40085
;
; LENGTH: 55
;
; TYPE: CDS
;

```

```

1 ORGANISM: HOMO SAPIENS
2
3 FEATURE:
4
5 1 OTHER INFORMATION: MAP TO AC004648.1
6
7 1 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
8
9 1 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.3

```

ON: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5  
ON: EXPRESSED IN PLACENTA, SIGNAL = 5.3  
ON: EXPRESSED IN HELA, SIGNAL = 4.8  
ON: EXPRESSED IN HEART, SIGNAL = 4.9  
ON: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
ON: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
ON: SWISSPROT HIT: P30999, EVALUE 3.00e-13  
ON: EST\_HUMAN HIT: A1936774.1, EVALUE 1.00e-22

4.4%; Score 280; DB 9; Length 55;  
urity 100.0%; Pred.No. 2.4e-11;  
nservative 0; Mismatches 0; Indels 0; Gaps 0;

THRSACGALNLVYGKANDNDKIALKNCGGIPALVRLRLKTTDLRELVLT 658

THRSACGALNLVYGKANDNDKIALKNCGGIPALVRLRLKTTDLRELVLT 55

Application US/09864761  
0048763AL

ON:

Sharron G.

el, David R.

el, David K.

l. Wensheng

ON: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

ON: GENE EXPRESSION ANALYSIS BY MICROARRAY

Aeomica-X-1

ION NUMBER: US/09/864,761

DATE: 2001-05-23

IN NUMBER: US 60/180,312

TE: 2000-02-04

IN NUMBER: US 60/207,456

TE: 2000-05-26

IN NUMBER: US 09/632,366

TE: 2000-08-03

IN NUMBER: GB 24263.6

IN NUMBER: US 60/236,359

TE: 2000-09-27

IN NUMBER: PCT/US01/00666

TE: 2001-01-30

IN NUMBER: PCT/US01/00667

TE: 2001-01-30

IN NUMBER: PCT/US01/00664

TE: 2001-01-30

IN NUMBER: PCT/US01/00669

TE: 2001-01-30

IN NUMBER: PCT/US01/00665

TE: 2001-01-30

IN NUMBER: PCT/US01/00668

TE: 2001-01-30

IN NUMBER: PCT/US01/00663

TE: 2001-01-30

IN NUMBER: PCT/US01/00662

TE: 2001-01-30

IN NUMBER: PCT/US01/00661

TE: 2001-01-30

IN NUMBER: PCT/US01/00670

TE: 2001-01-30

IN NUMBER: US 60/234,687

TE: 2000-09-21

IN NUMBER: US 09/608,408

TE: 2000-06-30

IN NUMBER: US 09/774,203

TE: 2001-01-29

ON NOS: 49117

ix Sequence Listing Engine vers. 1.1

TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005610.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3  
OTHER INFORMATION: SWISSPROT HIT: P30999, EVALUE 4.00e-08  
OTHER INFORMATION: EST\_HUMAN HIT: A126185.1, EVALUE 5.00e-17  
US-09-864-761-38977

Query Match 3.9%; Score 246; DB 9; Length 50;  
Best Local Similarity 100.0%; Pred.No. 3.7e-09;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; G

QY 880 WSVYIRAAVRKEKGLPILVELLRIDNDRVVCAVATARNMALDVRNKELI 929

Db 1 WSVYIRAAVRKEKGLPILVELLRIDNDRVVCAVATARNMALDVRNKELI 50

## RESULT 14

US-09-864-761-39938

Sequence 39938 Application US/09864761

Patent No. US20020048763AL

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID P

FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

ION NUMBER: US 09/774,203

ATE: 2001-01-29

ID NOS: 49117

max Sequence Listing Engine vers. 1.1

o sapiens

TION: MAP TO AC003089.1

TION: EXPRESSED IN BRAIN, SIGNAL = 1.8

TION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5

TION: EXPRESSED IN LUNG, SIGNAL = 1.4

TION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4

TION: EXPRESSED IN PLACENTA, SIGNAL = 1.7

TION: EXPRESSED IN HEART, SIGNAL = 1.4

TION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4

TION: SWISSPROT HIT: P30999, EVALUE 1.00e-09

TION: EST\_HUMAN HIT: BE177290.1, EVALUE 8.00e-18

38

larity 3.7%; Score 239; DB 9; Length 44;

larity 100.0%; Pred. No. 8.9e-09;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

WRDPPEVIMQLHQFPVSQSNAAAYLQHLCFGDNKIKAE 587

WRDPPEVIMQLHQFPVSQSNAAAYLQHLCFGDNKIKAE 44

8

Application US/10128714

US20030119013A1

TION:

ng, Bo

Wenqi

Shkoff, Daniel

udio, Carlos

oshkin, Alexey M

mieux, Sebastien M

TION: Identification of Essential Genes in Aspergillus fumigatus and

TION: Methods of Use

: 10182-018-999

ATION NUMBER: US/10128,714

DATE: 2002-04-23

ION NUMBER: US 60/285,697

ATE: 2001-04-23

ION NUMBER: US 60/287,066

ATE: 2001-04-27

ION NUMBER: US 60/295,890

ATE: 2001-06-05

ION NUMBER: US 60/303,899

ATE: 2001-07-09

ION NUMBER: US 60/316,362

ATE: 2001-08-31

ID NOS: 8603

ntin version 3.1

ergillus fumigatus

8

larity 3.6%; Score 227.5; DB 14; Length 1832;

larity 20.5%; Pred. No. 1.2e-05;

Conservative 181; Mismatches 503; Indels 485; Gaps 79;

SEKTSLSPLNGDSESTTSAILASVKEQLOPERLTREAEARQIVASQLE 81

QKRTSAEGPPITRKSTQVIDSINAVSDSMFSKQLP-----AAAEIN 319

KLGSETGSMSSMAEE-QFOWQSQ-----DGQKDIEDELT--GLELVDSCIRSLQ 132

Db 320 NSLAVETADGEIKNSPSEEDLAARWQAELEDDADETMTPTEDDLAARWQAELEDD--  
QY 133 ESGILDPPQDYSTGERP-----SILSQSALOLNSKPEGSFO-----YPASY-  
Db 374 DDDLLDDDDTTNAQRPEEAANIDHMDTSMLO---SPFGTPENLARPKVQPVSYT  
QY 177 TLALGETTSPQLPARGTOARATQGSFSGTTRAGHLAQPEPAPPPPPPPR-EPF  
Db 431 TSDL-----LSGIPAQNTAAQPTNASMSYFSAQ-----APPNVTTRAESF  
QY 232 PSLGSFAHLPD--APAAAAA---ALYSSSTILPAPRGSGPLAAPQGSPTKLO  
Db 478 EGYKSPYDIPEDLARPRRANSTVVAQGTVPKPPPRSSSIAP----PLKAS  
QY 287 PEGATVAAPRGSS-----PKQSPRLAKS---YSTSSPINIV-----  
Db 534 PLGTSSTATAQKNFEEELPLPPPRPKSRPASSGRYTNAPVASFSLQSIIPP  
QY 321 -VSSAGLSPIRVTSPPTVQS-----TISSSPIHQLSSTIGTYATLSP  
Db 594 NVFQAPQSNIGPPDPQLQOQPERLDPSNLLAPNVPAP--AVPSTASRY---SP  
QY 367 HASEQYKSHSQELYATATLQPGSLAAGSRASYSSOHGHLGPELRALQSPHHID  
Db 649 QAGVKPPSPR--YSPAPPQSTNAVAAPRNRYASQPASISGOGAALQFQRTSS  
QY 424 ---YEDRVYQKPPMRSLSQSQGDPLPAPHTGTYRTSTAPSSP-GVD---SVPLQ  
Db 707 EKIHYEDQGSERPOLQSTASPPPLNHSHPSEQPVSSSENKGFSGVDVLENVPL  
QY 476 HGPQNAATTFQASYA-----AGP-----ASNYADPPRLQYC  
Db 767 SPPKNPYAPSAVTNEFANRVAPVSTGPTAGMTGVLNLSSTESFPVPPRRSQTS  
QY 512 -----VESPYKSGPALP-----PEGTLARSPS-----  
Db 827 SPRLSVPSLDPPQRPASVHGTSPTTVPYAPAPVPTHNRAFSQVLEFPPTDG  
QY 538 IQKDPREFGWRDPEL-----PEVIQMLHQFPVSQSNAAAYLQHLCFGDNKIKAE  
Db 887 LER-----WKGAPIKFGGAGVISCFFPKHPIRYSAQAAPMKSCPG-----  
QY 592 GTQLLVLDLHDHMTVHRSCAGALNLYVGKAND-----DNKIALKNCGIIF  
Db 933 --ISQLNDWLPAEAGIVQHP--GPLK-----GKSKKDLVAVLSSKTAAPENANIF  
QY 645 LRKTDLTLEIRELTVGLWNL-----SSCDALK-----MPIIODALA-  
Db 985 SPDAS--KUREXT-LLWKVIRVLVENDGVLEGSVEAQKSLRNLLEPNLQDSGFN  
QY 681 VLT-NAVIIIPHSGWENSPLO-----DDRKIQLHSSQVLRNATGCLRNVSAGEAR  
Db 1042 VFTPSATLQ-----LNAPSQPDVDSKSVDL-----LRDTLVGEREK  
QY 736 CDGLTDALLYVIOALGSS-----EIDSKTVENCVCILRNLS--YRLAAETSC  
Db 1086 VDKRLWGHAMIIASRMDRSVMQVQVQFVRREVRSAFRTESLAAYEILLAGNIE  
QY 787 TDELDDLGCCEANGKDAESSGCGKKKKKKSQDQWDGVGPLPDCAEPK--GI  
Db 1141 -ESIDELV-----  
QY 844 PSIVKPYLTLSECSNPDTLEGAAGALQNLAAAGSWKWSVYIRAAVRK---EKGLP  
Db 1159 -----SKVDGHGPAKNSLDGL--DSNRETGLVLSNRSPPDQDALV  
QY 901 LRIDNRVVCATAALRNMA----LDVRNKELIGKYAMRDLVHRLPGGNSNNNTA  
Db 1203 LLSYNTEAAHICFIILSRVAVFGGLDDPOANIV-----LLGVQHRUSS  
QY 956 SDDTV-----TAVCC-----TLHEVITKMNENAKALRDAGGIEK-----

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

In search, using sw model

File 22, 2004, 12:10:45 ; Search time 27 Seconds  
(without alignments)

4364.245 Million cell updates/sec

-09-501-171A-4

77

MFARKPFGAALPLGAMPVDPQ.....PYSELNYETHYPASPDWV 1225

CSUM62

pop 10.0 , Gapext 0.5

3366 seqs, 96191526 residues

ts satisfying chosen parameters: 283366

3th: 0

3th: 2000000000

inimum Match 0%

Maximum Match 100%

listing first 45 summaries

IR 78:\*

pir1:\*

pir2:\*

pir3:\*

pir4:\*

the number of results predicted by chance to have a  
c than or equal to the score of the result being printed,  
ad by analysis of the total score distribution.

#### SUMMARIES

seq	Length	DB	ID	Description
5.1	1247	2	T42209	neural plakophilin
1.6	911	2	S28498	gene p120 protein
4.8	1050	2	T26395	hypothetical prote
1.2	725	2	S60712	band-6-protein - h
7.9	295	2	S60711	band-6-protein - b
4.0	1184	2	S50832	atrophin-1 - human
3.9	1184	2	G01763	atrophin-1 - human
3.6	1006	2	T42731	atrophin-1 related
3.5	3938	2	T42761	Bassoon protein -
3.4	528	2	T47141	gastric mucin (clo
3.4	1357	2	T29285	hypothetical prote
3.3	1560	2	T00080	hypothetical prote
3.3	2232	2	T34434	hypothetical prote
3.3	1198	2	T49726	hypothetical prote
3.2	806	2	JC4835	beta-catenin - Hyd
3.2	2649	2	T51023	hypothetical prote
3.2	1733	2	S27939	tensin - chicken
3.2	721	2	E70766	hypothetical prote
3.2	1487	1	EDBEF6	155K transcription
3.2	817	2	S51342	verprolin - yeast
3.2	1320	2	JC5630	TCOF1 protein - mo
3.1	2845	2	I49505	adenomatous polypo
3.1	839	2	T04859	extensin homolog F
3.1	580	2	T43481	probable mucin DKF
3.1	1151	2	T18535	high molecular mas
3.1	954	2	T19765	hypothetical prote
3.1	1851	2	T19964	hypothetical prote
3.1	1211	2	T42230	AF4 protein - mous
3.1	2715	2	T13049	eyelid - fruit fly

30 194 3.0 660 1 QBE3  
31 194 3.0 2843 1 RBHUAP  
32 193.5 3.0 1032 2 T34433  
33 193 3.0 1323 2 T30253  
34 192.5 3.0 3942 2 T42730  
35 191.5 3.0 781 2 A38573  
36 191.5 3.0 1896 1 RNFF2L  
37 190.5 3.0 1142 2 T00022  
38 190 3.0 878 2 T21621  
39 190 3.0 1487 1 EDBE1  
40 189.5 3.0 781 2 S35091  
41 189.5 3.0 1099 2 A56155  
42 188.5 3.0 1148 2 F86403  
43 188.5 3.0 1273 2 S58782  
44 188 2.9 1217 2 T42625  
45 188 2.9 1460 1 EDBE1F

BHLF1 prote  
adenomatous  
hypothetica  
salt prote  
Bassoon prc  
beta-cateni  
DNA-directe  
B120 protei  
hypothetica  
immediate-e  
beta-cateni  
tumor suppr  
probable tr  
SEC31 prote  
AF-4 protei  
immediate-e

#### ALIGNMENTS

##### RESULT 1

T42209

neural plakophilin related arm-repeat protein NPRAP - mouse

N;Alternate names: plakoglobin/armadillo protein

C;Species: Mus musculus (house mouse)

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec

C;Accession: T42209

R;Paffenholz, R.; Franke, W.W.

Differentiation 61, 293-304, 1997

A;Title: Identification and localization of a neurally expressed membe

A;Reference number: Z22075; MUID:98002299; PMID:9342840

A;Accession: T42209

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1247 <PAF>

A;Cross-references: EMBL:U90331; NID:g2580536; PID:g2580537; PIDN:AA8

A;Experimental source: brain

C;Genetics:

A;Gene: NPRAP

A;Note: predominantly, if not exclusively, expressed in neural and neu

Query Match 96.1%; Score 6129; DB 2; Length 1247;  
Best Local Similarity 94.8%; Pred. No. 1.3e-312;  
Matches 1185; Conservative 8; Mismatches 29; Indels 28; C

Qy	1	MEARKPGGAALPLGAMPVDPQPSASEKTSLSGLNTSNGDGSSETTTSALLASVK
Db	1	MEARKQSGAAPPFGAMPVDPQPPSASEKNSLSGLNTSNGDGSSETTTSALLASVK
Qy	61	QFERLTRELEBAERQIVASQLERCKLGSSETGSMSSMSAEQFQWQSQDQKDIIDE
Db	61	QFERLTRELEBAERQIVASQLERCKLGSSETGSMSSMSAEQFQWQSQDQKDIIDE
Qy	121	LELVDSICRSLQSGILDPODYSTERPSLLSOSALQNSKPEGSFOYPASYHSNQ
Db	121	LELVDSICRSLQSGILDPODYSTERPSLLSOSALQNSKPEGSFOYPASYHSNQ
Qy	181	GEITPSQLPARGTCARATGOSFSGCTTSRAGHLAGPEPAPPPPPPPPPPPPSLGS
Db	181	GDTPSQLPARSTQARAGSFSGCTTGRAGHLAGSEPA-PPPPPPPPPPPPPSLGS
Qy	241	PDAPPAALAAALYSSSTLPAPPRGSPPLAAPQGGSPKLRQGSAPAGATYAAPR
Db	240	PDAPP--AAALYSSSTLPAPPRGSPPLTTQGGSPKLRQGSAPAGAAVAAPR
Qy	301	KQPSRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVQSTISSPHQLSTIGTY
Db	298	KQPSRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVQSTISSPHQLSTIGTY
Qy	361	PTKRLVHASQYKHSQELVATATLQPGSLAAGSRASYSSQHGLGPELRLQSP
Db	358	PTKRLVHASQYKHSQELVATATLQPGSLAAGSRASYSSQHGLGPELRLQSP



[illegible]

QY	974	NWENAKALRDAGGIEKLVGISKSGDKHSFKVVKAAASQVLNSMWQYRDLRLSLYKKI
	:	: : : : : : : : : : : : : : :
Db	891	SAAATKDVBHKGTDKLRLRSRY-PTYSHRVCKYASQLYVMQHKLCHLHDGFKRK

[illegible]

```

C;Accession: S60712
R;Hatzfeld, M.; Kristjansson, G.I.; Plesgmann, U.; weber, K.
C;Date: 20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change 08-Oct-1996

```

A;Title: Band 6 protein, a major constituent of desmosomes from stratified squamous epithelium  
A;Reference number: S60683; MUID:95074299; PMID:7527055  
A;Accession: S60712

A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-725 <HAT>  
A;Cross-references: EMBL:X79293; NID:535014; PID:CAA558A1.1; PID:553

Query Match 11.2%; Score 713; DB 2; Length 725;  
Best Local Similarity 29.8%; Pred. No. 9.3e-30;  
Matches 218; Conservative 110; Mismatches 271; Indels 132; G

```
QY      QY          365 LVHASEQYSKHSDELATATLQPFGSLAG-----SRASYSSO----HGHL
         :::|||:|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      Db          46 MMTVRQKSXSOS--STLSHRGSGWMDGLADNNYGTTRSSYYSPQRAGNSGW
```

QY 411 RALQSP EHHIDPIYEDRVYQKPPMRSLSQSGDPLPPAHTGTYRTSTAPSPGVDS

471	RTGSHGPQAAAAAT-----PQRASYAAGPASN-YADPYRQLQY-----CPS
124	EWNRHYPGSCNNTGAGSDICFMQIKASGSEPLDYCDPGRGLKSGTGLSGKGTG

QY            516 YS-----KSGPALPPEGTLARSP-STDSTQKPREFGW-----RDP  
             ||     |     |     |     |     |     |     |  
Db            184 YSFYSTCGGAIAKKCFVRPPSCAKSQDVPVYTRHLLQDLFGHSRASGSKICSEDI

```

552 --LPEVIQMLQHQPSPVQSNAAAYLQHLCFGDNKTKAETIRQGGIQLLVLLDHRM
244 LTIKAVGYVLSODEKYOALGYVIOHTCFODESAKOOWOLGGTKIVALLRSPN

```

610	RSACGALRLNLYGKANDNDKIALKNGGIPALVRLRKTTDLLEIRELVTGLWNLS	QY
304	AAAGALRLNLFV--RSTTNKLETRRONGIRREAVSLRRTGNAE'OKOITGLWNLS	DB

QY	670	LKMP	IQAD	LAVL	TNAV	IIP	HSGW	ENSP	PLQD	DRK	TLQH	SSOV	LRN	ATG	CLRN	SSA
DB	362	LKEB	LIAD	PLAD	RVIT	IPF	SGW	CHG	NSM	REVV	---	DPEV	FNAT	GTG	CLRN	SSA

Qy	730	RRRR	CDGLT	DALLVI	QSLSS	SEDSKT	VENC	VCILR	NLSYR	LAETS	QGGHM
Dp	419	ROTWN	SGLTSD	MAVANC	AAAS	SPNDK	SVENC	VCVHNI	SVYR	DAEVD	TRYPK

Qy 790 LDGLLCGEANGKDA-----ESSGCGKGGKKKKKKSSQDQDWGVGLPDCAEPPKGIQML





C7	LARP--SIDSIQKDPREFGWRDEL-----PEVIOMLQHPSPVCSNAA 578
D5	PAGGPPLSATQIKOEPAE-EYETESPVPFARSPPPKKVDV-----PSHASQSA 759
C7	HLCFGDN 582
D5	HLDRGFN 770
PIEN	(man)
C7	#sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
D5	3
MBL	Data Library, March 1995
C7	G08343
D5	3
A7	ary; translated from GB/EMBL/DDBJ
A7	RNA
A7	<MAR>
A7	: EMBL:U23851; NID:G915325; PID:G915326
A7	B37
A7	: GDB:270336; OMIM:125370
A7	p-12p
	3.9%; Score 250; DB 2; Length 1184;
	arity 20.9%; Pred. NO. 3e-05;
	onservative 84; Mismatches 260; Indels 346; Gaps 42
C7	LNTSGNGSGSETTTTSAILASVKQELOFERLTRELEAERQIVASOLERECKLGSTG 90
D5	VTSISSDCK-----AEKSQTAKKARVEEASTPKVNKO-----G 71
C7	MSSAREQFOWSQDGQKDIEDEL-----TTGLELVDSICIRLSQESGILDPODYSTG 145
D5	ISESESE---ETWAPKTKTEELPRPOSPLDSLSDG--RSLNDGSSDPRDIQD 126
C7	-PILLSALQLNS-----KPEGSFQ----- 167
D5	SPSIYSGSVENDSDSSG--SQGPARYPHPPFPFPQPDPSTPROPEASFEPH 186
C7	ASYHSNQ---TLALGETTPS-----QLPARGTOARATQSFSQ---GTTSRAGHL 213
D5	TGYHAPMEPTSRMFQAPPACAPPHPOLYPGTGGVLGSPMPGPGKGGAASSVGCP 246
C7	PAPPP-----PPPPEPPAPSLGSAPHLPDAPPAANAALYYSSSTLP 260
D5	QHPPPTPTISVSSSGASGAPPTKPTTTPVGGG--NLPSAPP---PANFPHVTNPDL 301
C7	-----GGSP-----LAAPQ-----GGSPTKLQRGGSAPEGATYAAPR 296
D5	LRPLNNASAPPGGLGAQPLCHLPSHAMQGIGLPPGEKGTPLAPSLSLPPA 361
C7	KQSPRLAKSYSTSSPI-----NIIVSSAGLSPIRVTS-----PPTVQSTISSL 343
D5	PAPPMRPFYSSSSSSAAAASSSSSSSSASPFPASQALPSYHSPFPPTSLSVSN 421
C7	QLSSTTGTYATLS-----PTKRLVAS----- 369
D5	YTQPSLPQAWWSQGPPPPPYGRLLANSNAHPGFFPFTGAOSTAHPPPVSTHHH 481
C7	--EQYSKHSGELYATATLORPGS---LAAGSRASYSSQHG---LGFELRALQ-- 414
D5	QQQQQQQQQQHGNCGPPFPFGAFPHPLEGG----SSHIAHPYAMFPSLGLRAPY 536
C7	EH-----HIDIYEDRVYQKPMRSLSQ-----SQGD-----PLP 446
D5	AHLPPHSHQVSYSQAGPNGPSVVSSSSSSSSSTSGSYPCSHGPSQOGPAGYPFP 596



ABESTQTPSLTPSSDIPRISVGTSPMVQAQTGTHPRPSTP 1534

ne PGM-2A) - pig (fragment)  
ofa domestica (domestic pig)  
7 #sequence\_revision 21-Feb-1997 #text\_change 03-Nov-2000  
1: S55315  
askar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.  
06, 200, 1994  
ic mucin: isolation and characterization of a cDNA clone with a novel  
1 : I47141; MUID:94102478; PMID:7506218  
ary; translated from GB/EMBL/DDBJ  
RNA  
<TUB>  
askar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.  
9-96, 1995  
and characterization of cDNA clones encoding pig gastric mucin.  
1 : S55315; MUID:95275264; PMID:7555593  
5  
ary  
RNA  
<TU2>  
: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208  
submaxillary mucin  
3.4%; Score 217.5; DB 2; Length 528;  
arity 23.9%; Pred. No. 0.00048;  
onservative 73; Mismatches 255; Indels 95; Gaps 22;  
PD-----QPSSA-----LSPGLNTSGDGSETETTSAILASV 55  
|||||  
PPTTSVQPSGGSAPTTSATSVOTSSSPPISSTISVTSSSSVPPTTST--TSV 84  
LOFERLTREAEARQIVASQLERCKLGSETGSMSSMAEQFQMOSQDGKDIED 115  
SSSAPTTRATSVQ----SSSSSAPISSTTSVQPSGGSVPTTSATSVSSSSSA 140  
GLELVDSCTRLQESGILDPDYSTGERPSLLSQALQNSKPEGSFYQPAVSHN 175  
ATSVPSSSSPPISSTVSQPSSSSAPTTSATSV-----QPSSSSSPPIS--- 191  
LETTPSQLPARGTQARATQGSFSQCTTSRAGHLAGEPAEPAPPPPFRFPAPSIG 235  
VOTSSSSVPPTTTSVQPSSSSVPTTSATSV--VRSSSSSTPTTSTTSVQPSSS 249  
LPDAPFAAAALYYSSSTLPAPPFGGSLAAPQGGSPTKLQ--RGGSAP-EGATY 292  
---APTTSATSVQPSSSSTPIP-----SITSVQPSSSSAPTTSATSV 291  
SSS--PKGPSRLAKSYSTSSP-----INTIVSAGLSPIRVTSPPTVQSTISSSPI 346  
SSSSPPISTISVQPSSSSSSPTTSTTSVQPSSSGAP--TTTSATSVQPSSSSSP- 348  
STIG---IVATLSPTKRLVHASEOYXKHQELIYATATLQPCSLAAGSRASYSSH 403  
SITSVQPSSSSSSPTTST--TSVQPSSSGSAFTTSATSVQPSSSSVPTTSATSV-- 403  
PELRALQSPEHHIDPIVEDRVYQKPMPMSLSQSGDLPLPAHTGTVRTSTAPSSPG 463  
-----VRSSSSSSTPI-----PTTTSVQPSSSSVPTTSATSVQSTISSSSTPI 446  
PLQRTGSHGPQNAAAATFORASYAAGPASNYADPYRQLQCYCFSPVESYSKSGPAL 523  
SVQPSSSSSAP-TTSATSVQPSSSSSSPPISSTII---SVQPSSSSSSPTTSTTSVQ 501  
TLARSPIDSIQ 539  
3SATTSATSVQ 517

RESULT 11  
T29265  
hypothetical protein C01G8.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29265  
R:Du, Z.; Gattung, S.  
submitted to the EMBL Data Library, November 1996  
A:Description: The sequence of C. elegans cosmid C01G8.  
A:Reference number: Z20597  
A:Accession: T29265  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1357 <DUZ>  
A:Cross-references: EMBL:U80439; PIDN:AAB37645.1; GSPDB:GN00019; CESP:  
A:Experimental source: strain Bristol N2; clone C01G8  
C:Genetics:  
A:Gene: CESP:C01G8.7  
A:Map position: 1  
A:Introns: 89/1; 488/1; 701/3; 1056/2; 1159/3; 1197/1; 1312/3  
  
Query Match 3.4%; Score 217; DB 2; Length 1357;  
Best Local Similarity 18.0%; Pred. No. 0.0019;  
Matches 267; Conservative 162; Mismatches 518; Indels 540; G

QY 6 PGGAAPL-----GAMPV-PDQPSASSEKTSLSLSPG-----LNTE  
Db 96 PFGGPFQFGHGMDPNHYQQHGMPPHPGYPPOHMNAFSPGQYPGHQRP  
QY 43 SETETTSAILLASKEQLERLTRELEAEARQIVASQLERCKLGSETGSMSSMSF  
Db 156 GPFGGQAWEAPQHMQENDEHQRYAWHAQAAAAH-----HHF  
QY 103 QMOSQDGQKDIEDLTGLEVDSICRSLESGLDPQDYSTGERPSLLSQALQ  
Db 200 AAQAQAAQ-----AQQAATPAD---STTPAPAASSSSQAAE  
QY 161 -----KPEGSFQ--YPASVHSNOTL-----ALGETTSPQLPARG-----  
Db 237 NTOPATPFSASNQPTTFASASASTTLHVPEVASVAGGSQAGSRASFAGPGSATPDSE  
QY 193 TQARATQGSFSQCTTSRAGHLAGEPA-----PPP-----PPPREPFPAPSLG  
Db 297 ASAAADAQTTTSETSTPGPTSTSDTPAPSTSVPTTSVQPPQPPQPP  
QY 240 LPDAPFAAAALYYSSSTLPAPPFGGSLAAPQGGSPTKLQ---CGSAPEGATY  
Db 347 -PGQPPGP-----PGQQHPGYPGYGPGCPGAMPPAGFAPPGAPY  
QY 297 GSSPKOSPRLAKSYSTSSPINIVSAGLSPIRVTSPPTVQSTISSPHIULSSI  
Db 394 GAPP-----PAGEFHGHQHPQHAQ-----  
QY 357 ATLSPTKRLVHASEOYXKHQELIYATATLQPCSLAAGSRASYSSHGHGLGPELRA  
Db 414 -YLAWQQRYHQOQQHQOQQOQ-----GAPG-----GPRPPYPCGPVPP  
QY 417 EHHIDPIEDRVYQKPMPMSLSQSGDLPLPAHTGT-----YRTSTAPSSPGVDS  
Db 457 QNRMP-----PP-----PAQAPSPSAGSNKGKQPRYGTTPAPPSSRAPPT  
QY 470 QRTGSHGPQNAAAATFORASYAAGPASNYADPYRQLQCYCFSPVESYSKSGPALPP  
Db 503 SSTMPVAPSTSTQPTTPTTGSYL---ANTLATP-----GPAHAP  
QY 530 ARSPSIDSIQKOPREFGRWDRDPPEVIOMLQHFQPSVQSNAAAYLOHLCFGDNKIK  
Db 543 SHH-----HHVISQOQHYPGECIEATATSQ-----NQVK  
QY 590 ROGGIQLLDLLDHHRMTVEHRS-----ACGALENLYVGKAN-----

```
YARELINATPRRLIMSLRSLDAEAIWAINALNVLVYDDTNPHPTLQOMPLGVNVIV 633
-----DDNKIALKNC-----GGIPALVRL-----RKT 649
YATLSIMYPAEFQTEPKPIIMDESKEVEMKMAEKDGGIKSIIEKMPVVRKGS 693
IRELVTVLWNLSSCDALMKPIIODALAVLTVNAVIIHSGWENSFLQDRIQIHLSS 709
-----KTATFTWMSRNLQVHFDESIPALRKFRSREGSKNSDAMVDHSI---SS 743
--RNATGCLNRVNSAGEARRMRECDGLTALLY-----VIOQALGSSEIDS-- 758
EDRNSVGL---GGGLAEIATFLR-----DKLLHEKTRPRPVFSKYLLDDLETESTE 794
-----KTVENCVC----- 766
TVEVKKEIDEVAEKKDVLMRGAPDELSDCHEVELAWPRPTALSPKQAMEEL 854
-----TLRNI-SYRLAETSQ-----QHMGTDELGLLCEAN 799
ALALSNIILRGFSFVAGSDVLMARNEALLFIIGRLLLKLVNKEKISSKRPGIWSAGE 914
AESGCGWKKKKKKS-QDWDG-----VGPLPDCAEPPKGIQMLWHPISVKPY 850
APKQLSADKKRAKESVLDEVDATTAQMVETANQLRDDA-----FVMLTHMSVSLNL 969
LSECSNPDTLREGAALONLAAGSKWKSIVYIRAAVRKEKGLPILVELLIRIDNRVVC 910
-----PDAI-----APYIDGJLRWSV-SRVPEATDSSIPCPS-PRDYSLEIMC 1015
TALRNALDVNRKELIGKYA-MRLDVH---RLPGGNNNSNTASKA-----MSDD 958
VIERNVDMFLST---GWSRVEQFVHLITRLTWTNEETHYREFALVILNALCIASE 1071
AVCC-----TLHEVITKNWENAKALRD-----AG 985
YICAMETSIAIHLILFDSADQNQHVM--QAHGMAALRDNPMLGTSGVGLMRAAS 1129
KLVGISKSG--DKHSPKW-----KAASQVLNMMQYRDLRLSLYKKDQWSQ 1033
LLVKVPKAYKIYMKHQITLLOFTWSQLMDSRVAGMIADTLYEIOFFVKFEGED--- 1185
VASSSTIERDQRVSSSRTSISPRVSPN-NRSASAPASPRMISLKERKTYDEC 1092
-----VPEPLKITYSQYQPSQTDDEKSTNSDAAAPPQKSTSSSPKPVESAAAT 1236
NATYHGAKGEHTRSKDAMTAQNTGISTLYRNSY-GAPAEIDIKHQVSAQVPQPSR 1151
NVHNGNTLECTSSRPLSSEAAAAVITPYENHKNNEEEDLK-----KEIDEGESC 1290
ETYQPFQNSTRYNDESPEDQVHRRPPASRYTMHLGLKSTQNYV 1198
DT-BAIRLSQKSAALIDEQSRGPPPAKRAHLSNGFDKSKSV 1336

ein KIA0522 - human (fragment)
apiens (man)
99 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
80
akawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
', 1998
.on of the coding sequences of unidentified human genes. IX. The complet
r: Z14086; MUID:98290545; PMID:9628581
80
nary; translated from GB/EMBL/DBJ
mRNA
0 <NAB>
s: EMBL:AB011094; NID:g3043567; PIDN:BAA25448.1; PID:g3043568
urce: brain; clone HG1393
```

```
Query Match 3.3%; Score 213; DB 2; Length 1560;
Best Local Similarity 25.4%; Pred. No. 0.0038;
Matches 139; Conservative 54; Mismatches 147; Indels 208;

QY 48 TSAILASVKE-QBELFRLTRLEAERQIV---ASQLERCKLGETGSMSSMSI
Db 1140 TSLRESIAEVQEMEKYRVESELEKQKMMRPNASQPGGAK-DSVNGTM-ARSSI
QY 104 WQSQDQKQDIEDLTGL-ELVDSCIRSLQES-GILD-
Db 1197 -GAGDGLK--RGALSSSLRLSDAGKGRKNSVGSLSDTIEGVSISPRPHQME
QY 139 --PDYSTGRPSULLSOSALQLNS-----KPEGSFOYPASYHSNQTIALGETTPE
Db 1254 PPPEYKS-QRP--VNSSSSFLGSLFGSKGKGKGFQMPPP-----PTQASAS
QY 192 GTQ-----ARATGQSFQQTTSRAGHLAGEPAPPP-----
Db 1304 STHHHHHHHHSHGSHGGLVLPDQSKLOAL--HAQYCGQFGPAPPYLPPOQPE
QY 227 RE-PFAPSLGSAFHLDPDAPPAALAAALYYSSSTLPAPPRG-----GSPLAAPC
Db 1362 QQPPPLPQLGS---IP--PPAS-----APVGPHRHFAHGVPPGFC
QY 276 SPTKLQGGSAPEGATYAAPRGSSPKQSPRLAKSYSTSSPINIVVSSAGLSPIF
Db 1405 RPRAPRERGAG--GHQFAPHGRHPLHQT-----SPLI
QY 336 TVQSTSSSIHQLSSTIGTYATLSTKRLVHASEQYKSHQELYATATLQRPGE
Db 1441 -----APQHPAHKQG-----PKHFI-----FSHHPQMPAAGAAGGPGC
QY 396 RASYSSQHGHLGPELALQSPHHIDPIYEDRVVQKPPMRSLSQSQGDPLPPAHI
Db 1480 -GYSHPHP-----QSPLSHPFI-----PPHESY-----PPLPPPSI
QY 456 STAPSPGVDVSVLQRTGSHQGFQNAATAATFORASYAAGPASNYADPYROLQYCI
Db 1518 PLPTSP-----HGP-----LHAGGPPA-----OPTI
QY 516 Y-SKSGPA 522
Db 1545 RPSQAGSA 1552
```

## RESULT 13

```
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CERI
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 20
```

```
Query Match 3.3%; Score 211.5; DB 2; Length 2232;
Best Local Similarity 19.2%; Pred. No. 0.0074;
Matches 264; Conservative 160; Mismatches 575; Indels 379;

QY 7 PGAAPLGMPVPDPQSSASEKTSLSPLNTSGDGETTETTSAILASVKEQELC
```

```

STFASSTPIASSSPGSPVTVVAPG--SSSTYGSSTPSASSSSSGTMS-----T 412
AERQIVASQLERCKLKGSETGSMSSMAEEOFWQSDGQKDIEDLTLTGLGLVDS 126
TGSTVTVVAPVSSSTFGSSTPIASSSSG-----S 445
LOESGILDPQDYSTGERPRLISQALQNSKPEGSFYVPASYHNSQTLALGETTFS 186
V--SGSSSTYGSST--PSASSSA-GTASTISGSTGSTATIVPGSSSSSVGSSTQS 499
R-GTQARATGQSFQSGTTSRAGHLAGEPAPPPPPPPPPPPPP-----232
SPGTMSTVSGTGTSTTV-----VPGSSTSPASSSPNPSSPASTGTTITIGSS 554
-----SLGSAFHLDPAPAAAAAALYYSSTLAPPRGG-----266
STVSGSTVSGTGTSTQSTLASSTATPOSSSTVPSSSPQSSQSPAPNTGSTTSPQ 614
--SPLAAPQGSPTKLQGGSAPEGATYAAPRGSS-----PKOSP 304
SPSPKMPNPSSSTPTGSSQSTITPEGSTASSPTGSTGTSTFVATEVTSQSTVPSSG 674
KSYSTSSPINIVSSAGLSPIRVTSPTVOSTISS-----SPIHQLSS--351
STNSSPSPSLSLSPISQMSLT--TSEPSPSTQSSGAQSTLTTPSPNPSTSSLE 732
-----TICTYATLSPTKRLVHASEQYS-----373
3ATTSSGSGAGTTWT-SFQSSSVSSGSGSTSPAASITTSGMTSGSGTQTPGSSVST 791
-KHSQELIAT-----ATLQRP-----GSLAAGSRAYSYQHGHLPALQSP---416
JTSTQOSVSTNSPGSTVTRPSTVSGSTSGSTVTVGSTASTSGSSVASSSPAPST 851
DPIY---EDRVYQKP-PMRSISQSGDPLP-PAHTGTYRTSTAPS-----460
VPTSSGSSMITQSPVPSQSTSPVESSTTPSPGSPGTTLTSTSPSPSQSTTIGSTQ 911
PGVDSVPLQRT--GSGHGPQNAATPQRA-----SVAAGPASNVADPYRQLQCP 511
PGISTTSEMTSGSSTQTPGSTVTPQSTVSDSTSSGSTVTVGSTEGSSSIPS 971
-SPYSKGPAL---PBEGLTARPSIDSIQDPREFGWRDPPELPEVIMQLQHOFP 565
INPSTSGSSMSSTQTPQSSQSTSPVESSTSGATSSGSGPGTILTSI-----SPSP 1026
AAAYLQHLCPGDNKIAEIRROGGIQLLVLLDHRMTEVHSACGALR--NLVYK 623
[GS-----SQGSTSPVSTISQGSTETPGSTETPGSTVTKPSTVSGS 1069
YKI-----ALKNCGGIPALVRLLRKTTDLLEIRELVTGLVNLSSCDALKMPIIQ 677
TATMGSTASSTSGSSSTSPNSQSTSP-----STSGATSSP-----1111
JNAVILPHSGWENSLQDDRKQLHSSQVLRNATGCLRV---SSAGEARRMRRE 735
TLTSTISP-----SPSQ-----SSTIGSSQGSTSPVSTTSGDMSQSTQ 1156
DALLYVIOALGSSBIDSK---TVENCVCILR--NLSYRLAETSOGQHMGTDELD 791
GST--VTQPSVSGSTSGEITSQSTQTPRESSLSTSPALSTQSTQSSVSTNSPG 1214
EANGKDAESSGCW-----GKXKXKXKXQDQWDGVLPLDCAEPPKGI-----838
PSTVRGSTSGSTVTTGTEGSSSTSGSSSATLSLSSSPVPSSTSQSPNFTSGSST 1274
-----QMLWPSIVKPYLTLLSECSNPDTEGAAGALQNLAAG-----876
SOSTSPVSTTGTGEMTSHGSTQTP-STIGSTVTPSTVSGSSSGSTVTIGSSEA 1333
-SWKSVYIRAAVRKEKGLPILVELLRIDNDRVVCATALRNALMDVRNKELIGK 931
ISFKTSPSSISPVPTSSPIP-----STTFASSTSGSTISDVSSVST 1378

```

```

QY 932 YAMRDVLRHLPGG-----NNSNNTASKAMSDDTVTAVCTLHEVITKMNENAI
Db 1379 TSLAPLSSSLTPVPSSTQSFSSSEKSSASSPVPSTQSTSTPNP-TGSTESS
QY 984 A-GGIEKLVGTSKSGDKSPKVVKAASQVLNSMWQYRDLSLYKKDQWSQVHFVI
Db 1438 TISGSTQHTTMSKASSGSTSPSTNSQSTVTM-----GSSSTSGVE
QY 1043 IERDRQPYPS--SRTPSISPVRVSPNNRSASAPASPREMIS-----LKERP
Db 1484 SSTQPMSTSGSSAGSTVASSTASP-AASSTAPSSSTGTMSTSGSTVGTISESE
QY 1092 C---TGSNATVHGAKGHEHTSPKDAWTAQNTGISTLYRNSYCAPAEDIKHNOVSQAF
Db 1543 ASSQTGSTVTWSSSTSGVSTSSASTQ-PQMSTSGSSAGSTVASSTAGIVSTSI

```

RESULT 14  
T49726  
hypothetical protein B23L21.390 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49726  
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holla  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: 225022  
A:Accession: T49726  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1198 <SCH>  
A:Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.390  
A:Experimental source: BAC clone B23L21; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B23L21.390  
A:Map position: 6  
A:Introns: 10/3; 188/3

Query Match 3.3%; Score 207.5; DB 2; Length 1198;  
Best Local Similarity 22.6%; Pred. No. 0.005;  
Matches 141; Conservative 69; Mismatches 245; Indels 169; G

```

QY 7 PGAAPLGAM---PVPDPQSSASEKTSLSPLNTS-NGDGETETTTSA-----ILAS
Db 367 PGGVELGSIEKRPMP--PSRAGTFATQFSGASSTSLIGAAAEAPARQANDAPTVPK
QY 59 ELQFERLTRELEAERQIVASQ--LERCKLGSSETGSMSSMAEEOFWQSDGQKD
Db 425 NIPPTRTATLSQSSYGTASQSNFEPRLPGLGAAATGLSNYDAASSQSNFQPRPIR-
QY 117 LTTGLEAVDSICISLQESGILDQDYSTGERPSLLSQSALQNSKPEGSFYQYASY
Db 480 TATGISHVSS-VPSQQPS-----FQPRPLTRAATGMSNMGGP---SQSF
QY 177 TLALGE--TTPSQLPARGTQARATGQSFQSGTTSRAGHLAGEPAPAPPPP---PPPR
Db 525 PRELTRAATGLSNTSAGSQSQADFEPPRACTVPPQNFRFCMQQPPPPRAGTAPPR
QY 232 PSL---GSAFHLDPAPAAAAAALYYSSTLPA---PPRGSSPLAAPQGSSTKLQ
Db 585 PSMQPNSTF-----GPGYTDSPATYSTATMPAFQLPAR--APTAPMSYSNNNY-
QY 286 APEGATYAAPRGSSPKQSPSRKLSKYSTSSPINIVSSAGLSPIRVTSPTVOSTI:
Db 633 -----RGVEPR--PIERAYTGNVGNQSHDNFSAG-----
QY 346 IHQLSSTIGTYATLSPTKRLVHASEQYSK-----HSQELYATATLQRPGLSAAAGSI
Db 661 -NQVPTNNPYGGFSG-----HMSNCYDDGCGGYGQDIYGQDNYGQOQYMYNGK
QY 400 SQQHGLG-----PELRLQSPHEHI-----DPIYEDI
Db 712 HDNYGHDGYGPAQNNSQMPRLGSAAPPQTQSGYQPYTPQRTMTTAPPEGQPYHHK

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

in search, using sw model

ril 22, 2004, 12:06:54 ; Search time 19 Seconds  
(without alignments)  
3357.155 Million cell updates/sec

-09-501-171A-4  
77  
MEARKPPGAAPLGMPPVQ.....PYSELNYETSHYPASPDWV 1225

OSUM62

pop 10.0 , Gapext 0.5

1681 seqs, 52070155 residues

ts satisfying chosen parameters: 141681

gth: 0

gth: 2000000000

inimum Match 0%

aximum Match 100%

isting first 45 summaries

wiseProt\_42:\*

the number of results predicted by chance to have a  
r than or equal to the score of the result being printed,  
ed by analysis of the total score distribution.

SUMMARIES

ery	Length	DB	ID	Description
9.5	1225	1	CTD2 HUMAN	Q9ubq3 homo sapien
6.1	1247	1	CTD2 MOUSE	Q35927 mus musculu
4.2	1211	1	PKP4 HUMAN	Q99569 homo sapien
2.3	969	1	ARVC MOUSE	P98203 mus musculu
2.2	962	1	ARVC HUMAN	O00192 homo sapien
1.6	911	1	CTD1 MOUSE	P30999 mus musculu
1.5	264	1	CTD2 RAT	Q35116 rattus norv
1.5	968	1	CTD1 HUMAN	O80716 homo sapien
2.6	881	1	PKP2 HUMAN	Q99559 homo sapien
2.1	797	1	PKP3 HUMAN	Q9y446 homo sapien
2.0	797	1	PKP3 MOUSE	Q9gy23 mus musculu
1.1	728	1	PKP1 MOUSE	P97350 mus musculu
1.0	747	1	PKP1 HUMAN	Q13835 homo sapien
0.8	727	1	PKP1 BOVIN	Q28161 bos taurus
4.0	1185	1	DRPL HUMAN	P54258 homo sapien
3.6	1183	1	DRPL RAT	P54258 rattus norv
3.5	1324	1	SAL1 HUMAN	Q03173 mus musculu
3.5	802	1	ENAH MOUSE	Q03173 mus musculu
3.5	1300	1	SAL3 HUMAN	Q9bxa9 homo sapien
3.2	721	1	YK82 MYCTU	Q10690 mycobacteri
3.2	5560	1	SPEN DROME	Q8x83 drosophila
3.2	1322	1	SAL1 MOUSE	Q9gy74 mus musculu
3.2	5038	1	PCLO MOUSE	Q9gyx7 mus musculu
3.2	1487	1	ICP4 HSEB	P28925 equine heip
3.2	817	1	VRP1 YEAST	P37370 saccharomyc
3.1	1902	1	SNF1 HUMAN	O4497 homo sapien
3.1	2845	1	APC MOUSE	Q61315 mus musculu
3.1	1461	1	IE18 PRVIF	P11675 pseudorabie
3.1	741	1	GTSE MOUSE	Q8r080 mus musculu
3.1	1140	1	YM96 YEAST	O04893 saccharomyc
3.1	2716	1	OSA DROME	Q6in94 drosophila
3.1	2842	1	APC RAT	P70478 rattus norv
3.0	660	1	YHL1_EBV	P03181 epstein-bar

34	194	3.0	1575	1	SVJ1 HUMAN	O43426 homc
35	194	3.0	2843	1	APC HUMAN	P25054 homc
36	193.5	3.0	781	1	CTNB RAT	Q9wu82 ratl
37	193	3.0	1323	1	SAL3 MOUSE	Q62255 mus
38	192.5	3.0	4911	1	MLL3 HUMAN	Q8neza4 homc
39	191.5	3.0	781	1	CTNB HUMAN	P35222 homc
40	191.5	3.0	1083	1	T2D3 HUMAN	O00268 homc
41	190.5	3.0	1744	1	TENS_CHICK	Q04205 gal.
42	190	3.0	1087	1	ASPI MOUSE	Q62415 mus
43	190	3.0	1487	1	ICP4_HSEB	P17473 equi
44	189.5	3.0	781	1	CTNB MOUSE	Q02248 mus
45	188.5	3.0	1273	1	WEB1 YEAST	P38968 sacc

ALIGNMENTS

RESULT 1  
CTD2 HUMAN  
ID\_CTD2 HUMAN STANDARD; PRT: 1225 AA  
AC Q9UCB3; C00379; O15390; O43206; O43840; Q13589; Q9UM66; Q9UPM3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Catenin delta-2 (Delta-catenin) (Neural plakophilin-related ARM-1  
protein) (NPRAP) (Neurojuncin) (CT24).  
GN CTNND2 OR NPRAP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH PRESENILIN 1  
MEDLINE=97366296; PubMed=9223106;  
RA Zhou J., Liyanage U., Medina M., Ho C., Simmons A.D., Lovett M.,  
RA Kosik K.S.;  
RT "Presenilin 1 interaction in the brain with a novel member of the  
RT Armadillo family."  
RL NeuroReport 8:2085-2090(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.  
RC TISSUE=Brain;  
RX MEDLINE=99268159; PubMed=9971746;  
RA Lu Q., Paredes M., Medina M., Zhou J., Cavallo R., Peifer M.,  
RA Orecchio L., Kosik K.S.;  
RT "Delta-catenin, an adhesive junction-associated protein which pre  
RT cell scattering."  
RL J. Cell Biol. 144:519-532(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH PRESENILIN 1  
RC TISSUE=Fetal brain;  
RX MEDLINE=9923289; PubMed=10208590;  
RA Tanahashi H., Tabira T.;  
RT "Isolation of human delta-catenin and its binding specificity wit  
RT presenilin 1."  
RL NeuroReport 10:563-568(1999).  
RN [4]  
RP SEQUENCE OF 186-1225 FROM N.A. (ISOFORM 1), AND INTERACTION WITH  
RP PRESENILIN 1.  
RC TISSUE=Brain;  
RX MEDLINE=99155075; PubMed=10037471;  
RA Levesque G., Yu G., Nishimura M., Zhang D.M., Levesque L., Yu H.,  
RA Xu D., Liang Y., Rogava E., Ikeda M., Dutchie M., Murgolo N., Wan  
RA VanderVere P., Bayne M.L., Strader C.D., Rommens J.M., Fraser P.E  
RA St George-Hyslop P.H.;  
RT "Presenilins interact with armadillo proteins including neural-  
RT specific plakophilin-related protein and beta-catenin."  
RL J. Neurochem. 72:999-1008(1999).  
RN [5]  
RP SEQUENCE OF 350-1225 FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLIC  
RC TISSUE=Fetal brain;  
RX MEDLINE=98002299; PubMed=9342840;  
RA Paffenholz R., Franke W.W.;





MGTTDELGLLCEANGKDAESSGCGKKKKKKKKQDQNDGVGFLPDCAEPPKGIOM 840

SIYKPYITLLSECSNPDTLEGAAGALQNLAAAGSMKWSYVYIAAVRKEKGLPILVEL 900

SIYKPYITLLSECSNPDTLEGAAGALQNLAAAGSMKWSYVYIAAVRKEKGRPILVEL 900

NDRVCAVATLRNALDVRNKELIGKYAMDVLVHRLPGGNSNNNTASKMSDDTV 960

NDRVCAVATLRNALDVRNKELIGKYAMDVLVHRLPGGNSNNNTASKMSDDTV 960

CTLHEVITKNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAASQVLNSMWQYR 1020

CTLHEVITKNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAASQVLNSMWQYR 1020

LYKDKGWSQYHFVASSSTIERDRORPYSSTPTSPISPRVSPNNNSASAPAPREM 1080

LYKDKGWSQYHFVASSSTIERDRORPYSSTPTSPISPRVSPNNNSASAPAPREM 1080

ERKTDYECTSGNATYHGAKGHTSRKDAANTQNTIGISTLYXNSYGAPAEEDIKHNOV 1140

ERKTDYECTSGNATYHGAKGHTSRKDAANTQNTIGISTLYXNSYGAPAEEDIKHNOV 1140

VPOEPSRKDYETQYFPFONSTRNYDESFEDOVHHRPPASEYTMHLGLKSTGNYVDF 1200

VPOEPSRKDYETQYFPFONSTRNYDESFEDOVHHRPPASEYTMHLGLKSTGNYVDF 1200

RPYSELNYETSHYPASPSDWV 1225

RPYSELNYETSHYPASPSDWV 1225

STANDARD; PRT; 1247 AA.

Rel. 41, Created

Rel. 41, Last sequence update

Rel. 42, Last annotation update

-2 (Neural plakophilin-related ARM-repeat protein)  
(Jungin).

ND2 OR NPRAP.  
(Mouse).

tazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
heria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
390.

N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.

crain;  
299; PubMed=9342840;  
Frankie W.W.;  
on and localization of a neurally expressed member of the  
madillo multigene family."  
on 61:293-304(1997).

INTERACTION WITH E-CADHERIN.

158; PubMed=9971746;  
s M.; Medina M., Zhou J., Cavallo R., Peifer M.,  
Kosik K.S.;  
n, an adhesive junction-associated protein which promotes  
ag."  
144:519-532(1999).

SPLICING (ISOFORMS 1 AND 2), AND INDUCTION.

270; PubMed=10626844;  
Fan Q.W., Hayashi H., Michikawa M., Yanagisawa K.,  
f the mRNA for two isoforms of neural plakophilin-related  
protein/delta-catenin in rodent neurons and glial cells."  
c. 277:185-188(1999).

ICITY, AND SUBCELLULAR LOCATION.

926; PubMed=10753311;

```

247 AA; 134998 MW; D4A7A6B6A27D2919 CRC64;
/FTId=VSP_006747.
96.1%; Score 6129; DB 1; Length 1247;
larity 94.1%; Pred. No. 2e-283;
Conservative 8; Mismatches 29; Indels 28; Gaps 3;

RKPGAAPLAMPVDPDQSSASEKTSLSPLGINTSGDSETEITSAIASVKEQEL 60
RKQSGAAPFGAMPVDPDQSPSASEKNSLSPLGINTSGDSETEITSAIASVKEQEL 60
RLTRELEAERQIVASQLERCKLGETSGSMSSSAEEOFWQSGDQKQIDELTTG 120
RLTRELEAERQIVASQLERCKLGETSGSMSSISAGEQFHWQTDQKQIDELTTG 120
VDCIRSLOESGILDPQDYTGCRPSLLSQSALQNSKPEGSFOYPASYHSNQTLAL 180
VDCIRSLOESGILDPQDYTS*SERPSLLSQSALQNSKPEGSFOYPASYHSNQTLAL 180
TPSQLPARGTOARATGQSFSGTTSRAGHLAGPEPAPPPPPPPPPPPPPPSLGSFHL 240
APSQLPARGTOARAGQSFSGTTSRAGHLAGSEPA-PPPPPPPPPPPPPPPSLGSFHL 239
PPAAAAALYYSSSTLPAPPPGGGSLAAPQCGSPKLRQGSAPGEGATYAAPRGSSP 300
PP--AAAALYYSSSTLPAPPPGGGSLTTTQCGSPKLRQGSAPGEGATYAAPRGSSP 297
PSRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSTIGTYATLS 360
PSRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSTIGTYATLS 357
RLVHASQYKSHSELVATATLQRPGLAAGSRASYSSQHGHLGPELRALQSPHHI 420
RLVHASQYKSHSELVATATLQRPGLAAGSRASYSSQHGHLGPELRALQSPHHI 417
YEDRVYQKPPMRSLSSQSGDPLPAHTGTYRTSTAPSPGVDSVPLQRTGSHQPN 480
YEDRVYQKPPMRSLSSQSGDPLPAHTGTFTSTAPSPGVDSVPLQRTGSHQPN 477
ATFORASYAAGPASNADPYRQLYCPSPVESPYKSGPALPPECTLARSPSIDSIQK 540
ATFORASYAAGPASNADPYRQLYCASVDSPYKSGPALPPECTLARSPSIDSIQK 537
EFGWRDELPEVIQMLQHPFVSQSNAAAYLQHLQFQDNKIKABIRQGGIQLLVLD 600
EFGWRDELPEVIQMLQHPFVSQSNAAAYLQHLQFQDNKIKABIRQGGIQLLVLD 597
RMTEVHRSACGALRNLYGKANDDNKIALKNCGGIPALVRLRKTTOLEIRELVTGV 660
RMTEVHRSACGALRNLYGKANDDNKIALKNCGGIPALVRLRKTTOLEIRELVTGV 657
LSSCDALKMPIIQDALVLTNAVITPHSGWENSPLODDRKIQLHSSQVLRNATGCLR 720
LSSCDALKMPIIQDALVLTNAVITPHSGWENSPLODDRKIQLHSSQVLRNATGCLR 717
SAGEARRRMECCDGLTALLYVTSQALGSEIDSKTVENCVCILNLSYRLAETS 780
SAGEARRRMECCDGLTALLYVTSQALGSEIDSKTVENCVCILNLSYRLAETS 777
HMGTDGLDGLLCGFRANGKDAESSCGWKKKKKKKQDQWGVGVLPCAPPPKGIOM 840
HMGTDGLDGLLCGFRANGKDAESSCGWKKKKKKKQDQWGVGVLPCAPPPKGIOM 837
PSIVKPYLTLLSECSNPDTEGAAGALQNLAAAGSWK----- 879
PSIVKPYLTLLSECSNPDTEGAAGALQNLAAAGSWKGAEDVAGMAYALRSLPEGAP 897
-WSYVIRAAVKEKGLPLVELLRIDNDRVVCAVATARNMALDVRNKLIGKYAMR 935
QMSYVIRAAVKEKGLPLVELLRIDNDRVVCAVATARNMALDVRNKLIGKYAMR 957
HRLPGGNNNNNTASKMSDDTTAVCTTLEHVIITKNMENAKALRDAGGIEKLVGISK 995

```

958 DLVHRLPGGNNNNNTASKMSDDTTAVCTTLEHVIITKNMENAKALRDAGGIEK

996 SKGDKSPKVVVKAASOVLSNMWQYRDLRSLYKKQWSQYHFVASSSTIERDRQRI

1018 SKGDKSPKVVVKAASOVLSNMWQYRDLRSLYKKQWSQYHFVASSSTIERDRQRI

1056 TPISPVVSPNNRSASAPASPREMISLKERKTDYECTGSGNATYHGAKGHEHTSRP

1078 TPISPVVSPNNRSASAPASPREMISLKERKTDYECTGSGNATYHGAKGHEHTSRP

1116 QNTGISTLYXNSYGAEEDIKHNOVSAQVPOPSKDYETQYQPFQNSTYNDK

1138 QNTGISTLYXNSYGAEEDIKQNOVSTQVPOPSKDYETQYQPFQNSTYNDK

1176 VHRPPASEYTMHLGLKSTGNYVDYFSAARPYSGLNLYETSHYPASPDQSV 1225

1198 VHRPPASEYTMHLGLKSTGNYVDYFSAARPYSGLNLYETSHYPASPDQSV 1247

RESULT 3

PKP4 HUMAN

ID PKP4 HUMAN STANDARD; PRT; 1211 AA.

AC Q99569;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE Plakophilin 4 (p0071).

GN PKP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), AND VARIANT SER-44

RC TISSUE=Frontal cortex.

RX MEDLINE=97092329; PubMed=8937994;

RA Hatzfeld M., Nacheheim C.;

RT "Cloning and characterization of a new armadillo family member,

RT associated with the junctional plaque: evidence for a subfamily

RT closely related proteins.";

RL J. Cell Sci. 109:2767-2778(1996).

CC -I- FUNCTION: May play a role in junctional plaques.

CC -I- SUBCELLULAR LOCATION: Colocalized with desmoplakin at desmos

CC junctional plaques in cultured epithelial cells.

CC -I- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=Long;

CC IsoId=Q99569-1; Sequence=Displayed;

CC Name=Short;

CC IsoId=Q99569-2; Sequence=VSP\_006737;

CC -I- SIMILARITY: Belongs to the beta-catenin family.

CC -I- SIMILARITY: Contains 9 ARM repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a col

CC between the Swiss Institute of Bioinformatics and the EMBL ou

CC the European Bioinformatics Institute. There are no restric

CC use by non-profit institutions as long as its content is i

CC modified and this statement is not removed. Usage by and for

CC entities requires a license agreement (See <http://www.isb-sib.ch>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X81889; CAAS7478.1; -.

DR Genew; HGNC:9026; PKP4.

DR MIM; 604276; -.

DR GO; GO:0005911; C:intercellular junction; TAS.

DR InterPro; IPR008938; ARM.

DR InterPro; IPR000225; Armadillo.

DR Pfam; PF00514; Armadillo\_seg; 4.

DR SMART; SM00185; ARM; 6.

DR PROSITE; PS0176; ARM\_REPEAT; 3.

KW Cell adhesion; Cytoskeleton; Structural protein; Repeat; Coiled

KW Alternative splicing; Polymorphism.

FT DOMAIN 36 70 COILED COIL (POTENTIAL).

Db	737	KTVENCVC	TLNLSVRLEFVQPARLLG	NELDLLCKGSPSKDSEPS	-CWGKKK
Qy	818	--SQDQDVG	FLPDCABPKGIQMLWHP	SIYVKPYLTLLSECSNPDTL	GAAGAL
Db	796	TPQEQDQDVG	PIPLGSKSPKGVEMLWHP	SVVKPYLTLLAESNPATL	EGSAGSL
Qy	876	GSWKSVV	IRAAVKEKGLPIIVELLRI	DNDRVCAVATLRNALDV	RNKKELIG
Db	856	SNWKEAAV	IRGRPKRGLPIIVELLRM	DNDRVSSGATLRNALDV	RNKKELIG
Qy	936	DLVHRLPG	GNNSNTKASMSDDTVAV	CTCLHEVITKMNENAKAL	RDAGGIEKLA
Db	916	DLVNLPG	GNPS-----VLSDTMA	ICALHEVITKMNENAKAL	ADSGGIEKLA
Qy	996	SKGDKHS	PKVVKAAASQVLNSMWO	YRDLRLSLYKKGQWSQYH	FVASSTIERDRORP
Db	971	GRGRSSLL	KVVKAAAQVNTLWQYR	DLRSIYKKDGWNQNHFI	ITPVTSLDRPFKS
Qy	1056	TPSISPV--	RVSPNNRSASAPSPRE	MSISLKERKTDYCTGSN	ATYHGAKGEHTSI
Db	1027	PSLSTNT	QQMSPIISVSGTSSP	ALLGIRDPRSEYDRTQP	PMQYTNISQGDATHI
Qy	1114	TAQNTG	ISTIRNSYGAPAE	D---IKHNOVSAQVPQ	EPSRKDYETVQFPQNSTI
Db	1084	YPGSKSP	PIYISYSPAREQNR	LHQOQLYS--QDSSNR	KFNDAIRYLLOSPY
Qy	1170	SFFEDQV	HRPPASEYTMHLGLK	STGNVYDFYSAARPY	SELYNLTSHYPASPDSWY
Db	1142	PFYFDRV	HF--PASTDYSTQVGL	KSTTNVYDFYSTKRP---	SYRAEQYPGSPDSWY

RESULT 4

ARVC\_MOUSE

ID	ARVC_MOUSE	STANDARD;	PRT;	969 AA.
AC	P98203;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Armadillo repeat protein deleted in velo-cardio-facial syndrome homolog (Fragment).			
DE	Arvcf.			
GN	Mus musculus (Mouse).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20512094; PubMed=11058098;			
RA	Kaufmann U, Zuppinger C, Waibler Z, Rudiger M, Urbich C., Martin B., Jockusch B.M., Eppenberger H., Starzinski-Powitz A.;			
RT	"The armadillo repeat region targets ARVCF to cadherin-based cell junctions.";			
RL	J. Cell Sci. 113:4121-4135(2000).			
CC	!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHEREN			
CC	JUNCTIONS (BY SIMILARITY).			
CC	!- SIMILARITY: Belongs to the beta-catenin family.			
CC	!- SIMILARITY: Contains 10 ARM repeats.			

This SWISS-PROT entry is copyright. It is produced through a coll. between the Swiss Institute of Bioinformatics and the EMBL out the European Bioinformatics Institute. There are no restriction use by non-profit institutions as long as its content is in modified and this statement is not removed. Usage by and for c entities requires a license agreement (See <http://www.isb-sib.ch/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AJ243418; -, NOT\_ANNOTATED\_CDS.

HSSP; P35222; IG3U.

MGD; MGI:109620; Arvcf.

GO; GO:0005737; C:cytoplasm; IDA.

GO; GO:0005634; C:nucleus; IDA.

GO; GO:0005886; C:plasma membrane; IDA.

Qy	734	RECDGLTDALLXVIQ	ALSGSSSEIDSKTV	ENCVCILRNLSYRLAAETSQOQHMTG
Dy	536	RECEGLVDALLHALQ	SAVGRKDTDNKS	VENCVCIWNLNSYHVHKEVPGADRYQE
Qy	792	GLLCGEANGKDAES	SGCWKKKKKKSOD	QWD---GVGFLPDCAEPPKGIOMLWI
Dy	596	GSTTSORRRKD--	DASCFGGKAKGK	DAEMDRNFDTLDLPRTEAAKGPELLY
Qy	849	PYLTLSECSNPDTL	EGAAGALQNLAA	GSWKSVYVIRAAVRKEKGLPILVELLRJ
Dy	654	LYLSLLTESNFTLE	AAAGALQNL	SAGNWTWATVIRATVKEKRGFLPVIVELIQ
Qy	909	VCAVATALRNMAVD	NRNELIKICYAM	RDLVHRLPGGNNSNTA--SKAMSDDTVI
Dy	714	VRAVALALRNLSL	DRNNKDLISYAM	TSLVRNV---RNAQAPAHPSAHLLEEDTV
Qy	967	LHEVITKNMENAKA	LRDAGGIEKLV	GISKSGDKHSPKVKVKAASQVLNSMWQYRI
Dy	771	IHEIVSDSLDNARS	LLQARGVPAL	VALVAS---SQSVREAKAASHVLQTVMSYKE
Qy	1027	KDGNSQYHFVASS	TIERRQRPVSS	RRTSPISPV--RVSPNNRSASAPASPRE
Dy	828	QRDGWTKSRF-Q	SASTAKPGKT	PSSGGFDSTLPLVDKSLDGEKSNTRDVI
Qy	1080	--MTSLKERTDY	ECTGSNATYH	GKGEHTSRKDAMTANQNTGISTLYRNSYGA
Dy	887	GYATVDRRERT	-----LQSD	ST-----GD-TSEKELLRVYGGV-----YCGE
RESULT 5				
ARVC HUMAN				
ID	ARVC HUMAN	STANDARD;	PRT;	962 AA.
AC	AC	000192;		
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DN	Armadillo repeat protein	deleted in velo-cardio-facial syndrome.		
GE	ARVCF.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
NCBI_TaxID=9605;				
NCBI	[1]			
SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).				
RP	MEDLINE=97271559; PubMed=9126485;			
RA	Sirotkin H., O'Donnell H., DasGupta R., Halford S., St Jore B.,			
RA	Puech A., Parimoo S., Morrow B., Skoultschi A., Weisman S.,			
RA	Scambler P., Kucherlapati R.;			
RT	Identification of a new human catenin gene family member (ARVCF			
RT	the region deleted in velo-cardio-facial syndrome."			
RL	Genomics 41:75-83(1997).			
CC	-!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERE			
CC	JUNCTIONS.			
CC	-!- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=Long;			
CC	Isoid=000192-1; Sequence=Displayed;			
CC	Name=Short;			
CC	Isoid=000192-2; Sequence=Vsp 006739;			
CC	HEART SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUD			
CC	TISSUE SPRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.			
CC	-!- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) ;			
CC	HEMIZYGOS IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS.			
CC	HEMIZYGOSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE			
CC	PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPEC			
CC	PHENOTYPES, INCLUDING CONOTRUNCAL HEART DEFECTS, CLEFT PALAT			
CC	FACIAL DYSMORPHOLOGY.			
CC	-!- SIMILARITY: Belongs to the beta-catenin family.			
CC	-!- SIMILARITY: Contains 10 ARM repeats.			
This SWISS-PROT entry is copyright. It is produced through a coll				
between the Swiss Institute of Bioinformatics and the EMBL out				
the European Bioinformatics Institute. There are no restriction				

profit institutions as long as its content is in no way  
this statement is not removed. Usage by and for commercial  
res a license agreement (See <http://www.isb-sib.ch/announce/>  
il to [license@isb-sib.ch](mailto:license@isb-sib.ch))

```

AAC51202.1; -.
IG3U.
18; ARVCF.
1;
1; C:intracellular; TAS.
1; P:cell adhesion; TAS.
1; P:development; TAS.
1008938; ARM.
1000225; Armadillo.
1; Armadillo_seg; 4.
1; ARM; 5.
1; ARM; 5.
1; 76; ARM REPEAT; 3.
1; Cytoskeleton; Structural protein; Repeat; Coiled coil;
1; elic.
1 8 46
1 7 623
1 8 611
1 8 387
1 10 429
1 10 429
1 3 467
1 8 508
1 6 565
1 5 622
1 6 686
1 9 738
1 9 781
1 2 826
1 1 69
MEDCNVHSAASILASVKEQEARFELTRALEQERRHVALQL
ERAQQQPMVSGMGSGCQPLPMAWQQLVL -> MPAELR
(in isoform short)
/FTID:VSP 006739.
AA; 104641 MW; 74A1814A022FF2B1 CRC64;

```

[illegible]

Db	312	ADERFAFPMW-TAPLAQ-----PERGSMGLDR-----
Qy	515	PYSKSGPALPEGTIARSPSIDIQKDPREFGWDRPELPIVMIQHOFPSVQSNKA
Db	339	-----LVRSFSDVSARKEPR---WRDPFLPEVLAMLRHPDVPVKANA
Qy	575	QHLCFGDNKKIKAEIRROGGIQLLVDLLDRHMTVEVHRSACGALRNLYGKANDDNKI
Db	383	QHLCFENEGVKRRVRQUGRLPLVLLDHPRAEVRRACGALRNLSYGR-DTDNKA
Qy	635	CGGTPALVRLLRKTTDLREITLVGVNLNSCDALKMPILIQDALAVLTNAVLIIPH
Db	442	CGGVPALVRLLRARDNEVELVTGTLNLSYEFELKMWIIDHGLQTLTHEVIVPH
Qy	695	NSPLQDDRKIQLHSSQVLRNATCLENVSSAGEARRRMSCDGIITALLAVITQSA
Db	502	REPNEDSKPRDAEWTVFKNTSCLRNSSDGAEARRLRCEGLVDALLHQAQSA
Qy	755	EIDSKTVENCVCILRNLSYRLAAETSSQGOHMGTD---ELDGLLLGCEANG---KDAE
Db	562	DTDNKSVCENCVMRNLSYHVHKEVP---CADRYQEAEPPGLSGAVGSRRRRD
Qy	809	WGCKKKKKK---SQDWDG-----VGPIPLDCABPPKGIQMLWPSIVKPYLTLLS
Db	617	FGGKKAKEEMPHQCKKDGEMDRNFDTLDPKTEAAKGFELLYQFVVRLYLSILT
Qy	860	PDTLEGAAGALQNLAAAGSKWSVIIRAAVRKEKGLPIVLVLLRLTDNRVVCAVATA
Db	677	ENTLEAAGALQNLASGNWWTYIRATVRKERGLPVLVELLQSETDKVVRVAIA
Qy	920	ALDVNRNELIKGYAMRDLVHRL-----PGNNNSNTASKAMSDDTTAVQCTLL
Db	737	SLDRRNKDLIGSYAMAEVLRNVRNAQAPRPGA-----CLEEDTVAVVANTLI
Qy	972	TWYENAKALRDAGGIEKLVGISKSGDKHPKVVKAAASVLNSMWQYRDLRSLYK
Db	788	SDSLNARSLLQARGVPALVALVAS---SQSVREAKAAASHVLQTVWSYKELRGTLQ'
Qy	1032	SOYHFVASS-----
Db	845	TKAPFQGAATAKPGKALSPGQFSDTLPLVDKSLSEKGTGSRDVIIPMDALPGDG'
Qy	1044	ERDRQYFSSRTPSIS---PVRVSPNNRSASAPASPREMISLXKERTD 1089
Db	905	DRRRPRGASSAGEASEKEPLKLPD-SRKAPPPGPSPAPVLRVDAVGD 952
RESULT 6		
CTDI_MOUSE		
AD	CTDI_MOUSE	STANDARD; PRT; 911 AA.
IC	P30599;	
DT	01-JUL-1993	(Rel. 26, Created)
DT	01-JUL-1993	(Rel. 26, Last sequence update)
DT	28-FEB-2003	(Rel. 41, Last annotation update)
DE	Catenin delta-1 (p120 catenin)	(p120(ctn)) (Cadherin-associated S: substrate) (CAS) (p120(cas)).
DE	CTNND1	OR CATNS.
GN	Mus musculus	(Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OC	NCBI_TaxId=10096;	
RP	(1)_	SEQUENCE FROM N.A.
RP	CTNND1	OR CATNS.
RC	STRAIN=Swiss;	
RC	MEDLINE=93096477; PubMed=1334250;	
RA	Reynolds A.B., Herbert L., Cleveland J.B., Berg S.T., Gaut J.R.;	
RT	"p120, a novel substrate of protein tyrosine kinase receptors and	
RT	p60-src, is related to cadherin-binding factors beta-catenin,	
RT	Plakoglobin and armadillo."	
RL	Oncogene 7:2439-2445(1992).	
RN	(2)	
RA	Reynolds A.B.;	



all motility and cell scattering in response to  
growth factor treatment (in vitro).  
BINDS TO E-CADHERIN AT A JUXTAMEMBRANE SITE WITHIN THE  
C DOMAIN. BINDS TO PRESENILIN-1 (BY SIMILARITY).  
AR LOCATION: Adherens junction (By similarity).  
(: Contains at least 3 ARM repeats.  
(: Belongs to the beta-catenin family.

DT entry is copyright. It is produced through a collaboration  
Swiss Institute of Bioinformatics and the EMBL outstation -  
Bioinformatics Institute. There are no restrictions on its  
profit institutions as long as its content is in no way  
this statement is not removed. Usage by and for commercial  
uses a license agreement (see <http://www.isb-sib.ch/announce/>  
ail to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

;; BAA23384.1; -;  
00938; ARM.  
00225; Armadillo.  
; Armadillo\_seg; 2.  
3; ARM; 4.  
176; ARM REPEAT; 2.  
; Cytoskeleton; Structural protein;  
protein; Repeat.  
1  
20 59 ARM 1.  
34 104 ARM 2.  
20 162 ARM 3.  
19 255 POLY-LYS.  
34 264  
1 AA; 28928 MW; 065A886B8D2F58E0 CRC64;

21.5%; Score 1372; DB 1; Length 264;  
arity 100.0%; Pred.No.1.1e-58;  
nservative 0; Mismatches 0; Indels 0; Gaps 0;

NSNAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALNLVYG 622

NSNAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALNLVYG 60

NKIALKNCGGIPALVRLKRTTDLREIVTGVNLNLSDDLKWPITQDALAVL 682

NKIALKNCGGIPALVRLKRTTDLREIVTGVNLNLSDDLKWPITQDALAVL 120

IPHSWENSPLODDRKIQLHSSQVLNATGCLRNVSAGAEARRMRRCGDLTDA 742

IPHSWENSPLODDRKIQLHSSQVLNATGCLRNVSAGAEARRMRRCGDLTDA 180

QSALGSSBIDSKTVENCVCILNLSYRLAAETSQGHMGTDLDELGLCGEANGKD 802

QSALGSSBIDSKTVENCVCILNLSYRLAAETSQGHMGTDLDELGLCGEANGKD 240

ICWKKKKKKKSDODWDGVG 826

ICWKKKKKKKSDQWDGVG 264

STANDARD; PRT; 968 AA.

; O60713; O60714; O60715; O60935; Q9UP71; Q9UP72;

rel. 40, Created)

rel. 40, Last sequence update)

rel. 42, Last annotation update)

-1 (p120 catenin) (p120(ctn)) (Cadherin-associated Src

IS) (p120(cas)).

0384.

(Human).

azoa; Chordata; Craniata; Vertebrata; Euteleostomi;

eria; Primates; Catarrhini; Hominidae; Homo.

06;

SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
TISSUE=Petal kidney;  
MEDLINE=98317528; PubMed=9653641;  
Keirsebilck A., Bonne S., Staes K., van Hengel J., Nolllet F.,  
Reynolds A., van Roy F.;  
RT "Molecular cloning of the human p120ctn catenin gene (CTNND1):  
RT expression of multiple alternatively spliced isoforms.";  
RT Genomics 50:129-146(1998).

[2]  
SEQUENCE FROM N.A. (ISOFORM 1AC).  
TISSUE=Brain;  
MEDLINE=97349984; PubMed=9205841;  
Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes.  
RT The complete sequences of 100 new cDNA clones from brain which ca  
code for large proteins in vitro.";  
RT DNA Res. 4:141-150(1997).

-1- FUNCTION: EFFICIENT TYROSINE KINASE SUBSTRATE IMPLICATED BOTH  
CELL TRANSFORMATION BY SRC AND IN LIGAND-INDUCED RECEPTOR  
SIGNALING THROUGH THE EGF, PDGF, CSF-1 AND EBBB2 RECEPTORS. T  
ASSOCIATION OF CATEININS TO CADHERINS PRODUCES A COMPLEX WHICH  
LINKED TO THE ACTIN FILAMENT NETWORK, AND WHICH SEEMS TO BE O  
PRIMARY IMPORTANCE FOR CADHERINS CELL-ADHESION PROPERTIES.

-1- SUBUNIT: BELONGS TO A MULTIPROTEIN CELL-CELL ADHESION COMPLEX  
ALSO CONTAINS E-CADHERIN, ALPHA-CATENIN, BETA-CATENIN, AND GA  
CATENIN. BINDS TO PRESENILIN 1 C-TERMINAL FRAGMENT AND MUTUAL  
COMPETES FOR E-CADHERIN.

-1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR IN PARTICULAR C  
-1- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=32;

Comment=Experimental confirmation may be lacking for some  
isoforms;

Name=1ABC;  
IsoId=O60716-1; Sequence=Displayed;

Name=1AB;  
IsoId=O60716-2; Sequence=VSP\_006743;

Name=1AC;  
IsoId=O60716-3; Sequence=VSP\_006745;

Name=1BC;  
IsoId=O60716-4; Sequence=VSP\_006744;

Name=1A;  
IsoId=O60716-5; Sequence=VSP\_006743;

Name=1B;  
IsoId=O60716-6; Sequence=VSP\_006743;

Name=1C;  
IsoId=O60716-7; Sequence=VSP\_006744;

Name=1;  
IsoId=O60716-8; Sequence=VSP\_006743;

Name=2ABC;  
IsoId=O60716-9; Sequence=VSP\_006740;

Name=2AB;  
IsoId=O60716-10; Sequence=VSP\_006740;

Name=2AC;  
IsoId=O60716-11; Sequence=VSP\_006740;

Name=2BC;  
IsoId=O60716-12; Sequence=VSP\_006744;

Name=2A;  
IsoId=O60716-13; Sequence=VSP\_006740;

Name=2B;  
IsoId=O60716-14; Sequence=VSP\_006740;

Name=2C;  
IsoId=O60716-15; Sequence=VSP\_006744;

Name=2;  
IsoId=O60716-16; Sequence=VSP\_006740;

Name=3ABC;  
IsoId=O60716-17; Sequence=VSP\_006741;

Name=3AB;  
IsoId=O60716-18; Sequence=VSP\_006741;

Name=3AC;  
IsoId=O60716-19; Sequence=VSP\_006741;

Name=3BC;  
IsoId=O60716-20; Sequence=VSP\_006745;





RT	"Desmosomal plakophilin 2 as a differentiation marker in normal
RT	malignant tissues.";
RL	Differentiation 64:277-290(1999).
CC	-I- FUNCTION: May play a role in junctional plaques.
CC	-I- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes
CC	-I- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=2; Synonyms=B;
CC	ISOID=Q99959-1; Sequence=Displayed;
CC	Name=1; Synonyms=A;
CC	ISOID=Q99959-2; Sequence=VSP_006736;
CC	TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND AT DESMOSOMAL PL
CC	IN SIMPLE AND STRATIFIED EPITHELIA AND IN NON-EPITHELIAL TISS
CC	SUCH AS MYOCARDIUM AND LYMPH NODE FOLLICLES. IN MOST STRATI
CC	EPITHELIA FOUND IN THE DESMOSOMES OF THE BASAL CELL LAYER AN
CC	SEEMS TO BE ABSENT FROM SUPRABASAL STRATA.
CC	-I- SIMILARITY: Belongs to the beta-catenin family.
CC	-I- SIMILARITY: Contains 8 ARM repeats.
CC	This SWISS-PROT entry is copyright. It is produced through a coll
CC	between the Swiss Institute of Bioinformatics and the EMBL out
CC	the European Bioinformatics Institute. There are no restriction
CC	use by non-profit institutions as long as its content is ir
CC	modified and this statement is not removed. Usage by and for c
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/">http://www.isb-sib.ch/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X97675; CRA66264.1; -;
DR	EMBL; X97675; CRA66265.1; -;
DR	Genew; HGNC:9024; PKP2.
DR	MIM; 602861; -;
DR	GO; GO:0030057; C:desmosome; NAS.
DR	GO; GO:0016021; C:integral to membrane; TAS.
DR	GO; GO:0005634; C:nucleus; NAS.
DR	GO; GO:0005886; C:plasma membrane; TAS.
DR	GO; GO:0016337; P:cell-cell adhesion; NAS.
DR	InterPro; IPR008938; ARM.
DR	InterPro; IPR000225; Armadillo.
DR	Pfam; PF00514; Armadillo_seg.4.
DR	SMART; SM00185; ARM; 4.
DR	PROSITE; PS50176; ARM_REPEAT.1.
KW	Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
KW	Repeat; Alternative splicing.
FT	REPEAT 341..383 ARM 1.
FT	REPEAT 385..424 ARM 2.
FT	REPEAT 427..467 ARM 3.
FT	REPEAT 571..616 ARM 4.
FT	REPEAT 671..711 ARM 5.
FT	REPEAT 719..758 ARM 6.
FT	REPEAT 763..804 ARM 7.
FT	REPEAT 807..849 ARM 8.
FT	VARSPLIC 460..503 Missing (in isoform 1).
FT	/FTID=VSP_006736.
FT	SEQUENCE 881 AA; 97398 MW; 947838B760EFD1D5D CRC64;
QY	Query Match 12.6%; Score 801; DB 1; Length 881;
Db	Best Local Similarity 28.4%; Pred. No. 6.3e-31;
QY	Matches 259; Conservative 134; Mismatches 323; Indels 196; G
QY	256 SSTLPAPRGGSPIA--APQGGSPTK-----LQGGSAPEG-----ATYA
Db	27 SSSIALPSEAKLKGSSRGCGQTVKSLRIQEQVQTLARKGRSSVNGNMLHRTSS
QY	297 -----GSSPKQSPSLAKSYSTSSPINIVSSNAGLSPIRVTSPTPTVQS
Db	87 VYNLHVENDFVGRSFPVTKTYDMLKAGTTAT----YEGRWGRGTAQYSQKSVEE.
QY	344 SPIHQLSSTTIGTYATLSPTK-RLHVASEQYSK-----HSOELVATATLQRP--
Db	143 HPLRRLRISPDS-----SPERAHYTHSDYQYQSRQAGTLLHQQSRRAALLVPRPY.
QY	391 LAGSRASYSQGHGGLPELRALQSPHHIDPIYEDRVYQKP-----PMRSLSSQSQ

```

VSRAAGTTTQRHFDTYHQYQ--HGSVSDTVFDSIPANPALLTYPRPQTSRSMGNLL 256
-----FLPAAHTCTYRTSTAPSSGVDVSPLORTGSHQGPQNAATAFQ 487
NYLTAGLTGVGVRLVLPQPT-QNRASSRSHWSSHFSTRTLREAGPSVAVDSSGR 315
YAAGPASNYADPYRQLOVCPVESPKSGPALPEGTTLARSPSIDSIQDKDPREFGW 547
LTVGQAA-----AGSGNLLTERSTFTDS-----QLGN 346
E--LPEVLQMIQ--HQPPSVQSNAAAYLOHLCFGDNKIKAEIRPQGGIQLLVLLDH 603
EMTLERAVSMLEADHMPSPRSIAAATIQHECFQKSEARKRVNQLRGILKLLQLVK 406
EVHRSACGALNLYVGKANDNDKIALKNCGGIPALVRLAKRTDLEIRELVT----- 658
DVQRAVCGALNLFV--EDNDNKLEVAELNGVPELLQVLQKTRDLETKKQITHTVYN 464
-----GVLNLSGCDALKMPIIODAL 679
RNGWPGAVAHACNPSTLGGGGGGRITRSGVRDQDPDQHGLLNLSNDKLNLMITEAL 524
TNVLIIPHSWENSPLODDRKIQ-LHSSQVLRNATGCLRNVSSAGEARRMRREC DG 738
TENIILIPSGW---PEGDYPKANGLLDFDIFYNVTGCLRNVSSAGADGRKAMRRC DG 581
ALAYVIOQSALGSSEIDSKTVENCILRNLSYLAETS-----QGHWMGTDE 789
SLVHYVGTIADYQDDKATENCVCILHNLNSYQLEAELPEKYSQNIYIQRNITQTD- 640
LLCGEANGKDAESSGCKGKKKKKQSDQWDGVPPLDPCABPPKGIQMLWHPISIVKP 849
-----NNK---SIGCFGSRKVK--EQYQDV--PMPEEKSNNPKGVWLWHSIVIRM 685
LLSECSNPDTLEGAGALQNLAAQSKWKSIVI-RAAVRKEKGLPIVVELLRIDNDRV 908
LIAKSVNYTQEAALQNLQNLTAGSGMPSTVAGTVQKESGLQHTKMLHVGDPVS 745
VATALRNALDVRNKLIGYAMRDLVHRLPGGNNSNNTASKAMSDDTVTAVCCTLH 968
AISLLRNLSRLSLQNEIAKETPLDLVSIIP-----DTPVSTDLLIETTASACVTLN 800
TKWENAKALRDAGGIEKLVIKSKDKXH-SPKVVAASQVLSNMWQYRDLRLSKY 1027
QNSYQNAARDLNTGGIQKIMAI--SAGDAYASKAKAASVLLYSYLWAHTHELHAYK 858
#SQYHFVAS 1039
FKKTDVNS 870
STANDARD; PRT; 797 AA.
(Ref. 41, Created)
(Ref. 41, Last sequence update)
(Ref. 42, Last annotation update)
3.
(Human).
atazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Cheria; Primates; Catarrhini; Hominidae; Homo.
606;
N.A.
Carcinoma;
2495; PubMed=10374265;
Langbein D., Praetzel S., Rode M., Rackwitz H.-R.,
3 - a novel cell-type-specific desmosomal plaque
ion 64:291-306(1999).

```

```

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99310799; PubMed=10381383;
RA Bonne S., van Hengel J., Nollet F., Kools P., van Roy F.;
RT "Plakophilin-3, a novel armadillo-like protein present in nuclei
RL desmosomes of epithelial cells.";
RN J. Cell Sci. 112:2265-2276(1999).
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388237; PubMed=12477932;
RA Strausberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk
RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanche
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinaki M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in junctional plaques.
CC -!- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes
CC -!- TISSUE SPECIFICITY: FOUND IN DESMOSOMES OF MOST SIMPLE AND
CC STRATIFIED EPITHELIA. NOT FOUND IN FORESKIN FIBROBLASTS AND
CC CELLS OF LYMPHATIC FOLLICLES NOT FOUND IN NON-EPITHELIAL
CC DESMOSOME-BEARING TISSUES.
CC -!- SIMILARITY: Belongs to the beta-catenin family.
CC -!- SIMILARITY: Contains 8 ARM repeats.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a col
CC between the Swiss Institute of Bioinformatics and the EMBL ou
CC the European Bioinformatics Institute. There are no restrictio
CC use by non-profit institutions as long as its content is i
CC modified and this statement is not removed. Usage by and for
CC entities requires a license agreement (See http://www.isb-sib.ch
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z98265; CAB44310.1; -.
DR EMBL; AF053719; AAF23050.1; -.
DR EMBL; BC000081; AAH00081.1; -.
DR Genew; HGNC:9025; PKP3.
DR MIM; 605561; -.
DR GO; GO:0005911; C:intercellular junction; TAS.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 2.
DR SMART; SM00185; ARM; 3.
DR PROSITE; PS50176; ARM_REPEAT; 1.
KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein
KW Repeat.
FT REPEAT 305 348 ARM 1.
FT REPEAT 351 390 ARM 2.
FT REPEAT 393 432 ARM 3.
FT REPEAT 449 487 ARM 4.
FT REPEAT 491 536 ARM 5.
FT REPEAT 596 637 ARM 6.
FT REPEAT 645 684 ARM 7.
FT REPEAT 689 730 ARM 8.
SQ SEQUENCE 797 AA; 87081 MW; D43C7E77FA805E7E CRC64;
Query Match 12.1%; Score 770.5; DB 1; Length 797;

```



---LPDVRGLDSTYG-----HRTLQRLSS-----GFDDIDLPSAVKYLMA 330  
 SVQNSAAAYLOHLFCGDNKIKAEIRROGGIOLLVLLDHRMTVEHRSACGALRNLYV 621  
 NLQVLGAAYIOHRCVSDAAAKQARSLOAVPRLVKLFNHANQEVORHATGAVENLIV 390  
 NDDNKIALKNCGGIPALVRLLRKTTDLBIRELVITGVLWNLSGCDALKMPITIDALAV 681  
 NVDNKLALVEENGIFELLRTLRQDD--ELRKNVTGILWNLSGSDHLKDLRLARDTLEQ 447  
 AVIIPHSWENSPLODDRKIKIOLHSS--QVLRNATGCLRNVSAGEEAPRRRRECGL 739  
 LVLSPGSGPGPPL-----IQONASEAEIFYNTGFLRNLSASQATKQKRECHGL 502  
 LLYVIQSALGSEISDKTENCVCILRNLSYRL-----AAETSQGHMGTDDELDG 792  
 LVTVINHALDVGKCEDKSVENAVCVLRNLSYRLYDEMPSPSALQRLGR--GRDWAG 560  
 GEANGKDAESSGCWGKKKKKXKXQDWDGVPDPDCAEPPKGIQMLMHPSTIVKPYLT 855  
 G-----EMVGCFTPOSRRRLRELFETADALTFAEVSKDPKGLWMLSPQIVGLYNR 613  
 ECS-NPDTLEGAAGALONLAAGSKWKS-VYTRAAVRKEKGPILVELLRIDNRVVC 910  
 RCELNRHTTEAAGALONITAGDRRWAGVLSRLALEQERIINLPDDRVRTADHNNLR 673  
 TALRNALDVENKELIGKYAMRDLVHRLPGGNNNNNTASKAMSDDTVTAVCCTLHEV 970  
 GLIRNLSRNARNKDEMSTKVVSVHLIEKLPG-----SVGEKCPAELVNIATVNLN 728  
 NMENAKALRDAGGTEKLVGISKSGDKHSPKVKVAASQVLNMSMQYRDLRSLYKKDG 1030  
 SPTAARDLLYFDGRLKLVFKKKRSDSPDSEKSSRAASLLANLWQYSLKHLRDFRAKG 788  
 YHFV 1037  
 EDLF 795  
 STANDARD; PRF: 728 AA.  
 (Rel. 4i, Created)  
 (Rel. 4i, Last sequence update)  
 (Rel. 4i, Last annotation update)  
 1.  
 (Mouse).  
 etazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Theria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 0090;  
 M N A.  
 /6; TISSUE=Skin;  
 Hunziker A.H., Franke W.W.;  
 EP-1996) to the EMBL/GenBank/DBJ databases.  
 S; SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY  
 TY).  
 LAR LOCATON: Nuclear and associated with desmosomes (By  
 ty).  
 TY: Belongs to the beta-catenin family.  
 TY: Contains 7 ARM repeats.  
 ROT entry is copyright. It is produced through a collaboration  
 Swiss Institute of Bioinformatics and the EMBL outstation -  
 Bioinformatics Institute. There are no restrictions on its  
 -profit institutions as long as its content is in no way  
 this statement is not removed. Usage by and for commercial  
 uires a license agreement (See <http://www.isb-sib.ch/announce/>  
 mail to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CAA69240.1;

YTVRLMTSQPQAKQYFNSMLNVLNCRNTAS--SPKAAEAARLLSLDMWAS 695

3LYKKDQWSQ 1033

3VLRQGGFDR 709

STANDARD; PRT; 747 AA.

1; Q15152;  
rel. 41, Created)  
rel. 41, Last sequence update)  
rel. 41, Last annotation update)  
(Band-6-protein) (B6P).

(Human).

taxon: Chordata; Craniata; Vertebrata; Euteleostomi;  
Phyla: Primates; Catarrhini; Hominoidea; Homo.

N.A. (ISOFORM 1).

199; PubMed=7527055;  
Kristjansson G.I., Plessmann U., Weber K.;  
in, a major constituent of desmosomes from stratified  
a novel member of the armadillo multigene family.";  
107:2259-2270(1994).

N.A. (ISOFORM 1).

1994) to the EMBL/GenBank/DBJ databases.

N.A. (ISOFORMS 1 AND 2).

1972; PubMed=9369526;  
ungbein L., Rode M., Praetzel S., Zimbelmann R.,

1a and 1b: widespread nuclear proteins recruited in  
relial cells as desmosomal plaque components.";  
98. 290:481-499(1997).

SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES.

AR LOCATION: Nuclear. Isoform 1 is also associated with  
3.

FE PRODUCTS:

1; native splicing; Named isoforms=2;

1; nonyma=B;

13835-1; Sequence=Displayed;

1; nonyma=A;

13835-2; Sequence=VSP\_006735;

3; CIFICITY: NUCLEAR ISOFORM IS WIDELY EXPRESSED. ISOFORM 1  
1; IN STRATIFIED SQUAMOUS, COMPLEX, GLANDULAR DUCT AND  
1; ITHELIA.

1; DEFECTS IN PKP1 ARE ASSOCIATED WITH ECTODERMAL

1; SKIN FRAGILITY SYNDROME.

1; Belongs to the beta-catenin family.

1; Contains 8 ARM repeats.

1; entry is copyright. It is produced through a collaboration  
1; Swiss Institute of Bioinformatics and the EMBL outstation -  
1; bioinformatics institute. There are no restrictions on its  
1; profit institutions as long as its content is in no way  
1; this statement is not removed. Usage by and for commercial  
1; a license agreement (See <http://www.isb-sib.ch/announce/>  
1; ail to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CAAS5981.1; -

CAAB4426.1; -

-; NOT ANNOTATED\_CDS.

CAAB4426.1; -

223; PKP1.

-

-

7; C:desmosome; NAS.

DR GO: 0005634; C:nucleus; NAS.

DR GO: 0019215; F:intermediate filament binding; NAS.

DR GO: 0030280; P:structural constituent of epidermis; NAS.

DR GO: 0007155; P:cell adhesion; NAS.

DR GO: 0007165; P:signal transduction; NAS.

DR InterPro: IPR008938; ARM.

DR InterPro: IPR000225; Armadillo.

DR Pfam: PF00514; Armadillo\_seg; 4.

DR SMART: SM00185; ARM; 5.

DR PROSITE: PS00176; ARM\_REPEAT; 3.

KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;  
KW Repeat; Alternative splicing.

FT REPEAT 234 274 ARM 1.

FT REPEAT 277 316 ARM 2.

FT REPEAT 319 359 ARM 3.

FT REPEAT 359 399 ARM 4.

FT REPEAT 439 484 ARM 5.

FT REPEAT 537 577 ARM 6.

FT REPEAT 585 624 ARM 7.

FT REPEAT 626 670 ARM 8.

FT REPEAT 673 713 ARM 8.

FT VARSPLIC 412 432

FT Missing (in isoform 1).

FT /FTID=VSP\_006735.

FT R -> G (IN REF. 1).

FT CONFLICT 154 154

FT CONFLICT 216 222

FT CONFLICT 462 462

FT CONFLICT 496 496

FT CONFLICT 506 506

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

FT CONFLICT 553 553

Query Match 11.0%; Score 699; DB 1; Length 747;

Best Local Similarity 29.0%; Pred. No. 3.5e-26;

Matches 219; Conservative 108; Mismatches 271; Indels 156; G

365 LVHASTQYKSHQELATATLQRPGLAAG-----SRASYSSQ-----HGHL

46 MMTVTKRQSKSQS--STLSHNRGMYDGLADNYNYGTTSSRSYKFGAGNSW

411 RALQSPHHIDPIEDRVYQKPPMSLSQSGDPLPPAHTGYRTSTAPSSPGVDS

101 -----YPIYNGTLKREPDRRFS-----S

471 RTGSHGPNQAAAT-----FORASYAAGPASYADPYRLQY-----CPS

124 ENWSRHYPRGSCMTTGAGSDICPMQKIKASRSPDLYCDPRGLRKGTIGSKGQKT

516 YS-----KSGPALPPEGLARSPI-----DSIQKDPREGW-----R

184 YSPYSTCGQKATKCFVRPPSCASKQDPVYIPPIPCNKD-LSFGHSRASSKICSE

552 ----LPEVIQMLQHPFVSQSNAAAYLQHLFCGDNKIKAEIRQGGIQLLVLDLH

243 SGTIPKAVQLSSQDEKYQAIGAYVIOHTCFODESAKQOVYQIGGICKLVLLRS

608 VHSACALRNLYGKANDDNKIALNCGGIPALVRLRLKRTDLEIRELVGLWN

303 VQAAAGALRNLYF--RSTTNKLETRQNGIREAVSLRLRTGNAETQKQLTGLWN

668 DALKMPIDALAVLNNAVIIIPHSCHENSPLQDRDKIQLHSSQVLRNATCIR---

361 DELKEELIADALVLRADRVIIIPFSGMCDGNSNMSREV--DPEVFNATCLRKRL

721 -----NVSSGEEARRRMRECDGLTDALLVVIQSLAGSSEIDSITVEI

419 LLALVFORATSSRVNLSSA-DAGRTWRYNYSGLIDSLMAYVQNCVAASRCDDKVEI

767 ILENLSYRLAAETSSQGHMGTDELGLLCEANGKDA-----ESSGCGWKKKKKKKS

478 VLHNLVRLDAEVPTRYQ-----LEYNARNAYTEKSTGCFNSKSDKMMN

823 DGVGPLPDCAEPPKGIQMLWHPISIVPKYLTLLSECSNPDTLEGAAGALQNLASGW

527 D--CPLPEETNPKGSGWLYHSDAIRTYLNLGMGSKKDKATLEACAGALQNLATSKG



```
CC CC -!- CAUTION: Ref.2 sequence differs from that shown due to severe  
CC frame-shifts.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a col-  
CC lation between the Swiss Institute of Bioinformatics and the EMBL out-  
CC put by the European Bioinformatics Institute. There are no restriction  
CC use by non-profit institutions as long as its content is irrevocably  
CC modified and this statement is not removed. Usage by individuals or  
CC entities requires a license agreement (See http://www.isb-sib.ch/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; D31840; BAA06626.1; --  
DR EMBL; D38529; BAA07534.1; ALT_FRAME.  
DR EMBL; U23851; AAB50276.1; --  
DR EMBL; U47924; AAB51321.1; --  
DR EMBL; D63808; BAA23631.1; --  
DR EMBL; L10377; -: NOT_ANNOTATED_CDS.  
DR PIR; G01763; G01763_-DEPLA.  
DR Genew; HGNC:3033; DEPLA.  
DR MIM; 607462; --  
DR MIM; 125370; --  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0005634; C:nucleus; TAS.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR GO; GO:0007417; P:central nervous system development; TAS.  
DR InterPro; IPRO02951; Atrophin.  
DR Pfam; PF03154; Atrophin-1; 2.  
DR PRINTS; PR01222; ATROPHIN.  
KW Triplet repeat expansion; Polymorphism; Epilepsy.  
FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE)  
FT DOMAIN 302 305 POLY-PRO.  
FT DOMAIN 376 382 POLY-SER.  
FT DOMAIN 386 397 POLY-SER.  
FT DOMAIN 442 447 POLY-HIS.  
FT DOMAIN 479 483 POLY-GLN.  
FT DOMAIN 484 497 POLY-PRO.  
FT DOMAIN 504 507 POLY-SER.  
FT DOMAIN 564 574 POLY-PRO.  
FT DOMAIN 704 707 ARG/ALA-RICH (MIXED CHARGE)  
FT DOMAIN 802 815 ARG/GLU-RICH (MIXED CHARGE)  
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE)  
FT DOMAIN 925 934 MISSING (IN REF. 3).  
FT CONFLICT 94 94 H -> Y (IN REF. 1).  
FT CONFLICT 333 333 M -> T (IN REF. 3).  
FT CONFLICT 339 339 P -> I (IN REF. 6).  
FT CONFLICT 541 541 A -> G (IN REF. 1).  
FT CONFLICT 1028 1028  
SQ SEQUENCE 1185 AA; 124773 MW; 5135744CEE491C18 CRC64;  
  
Query Match 4.0%; Score 255.5; DB 1; Length 1185;  
Best Local Similarity 20.3%; Pred.No. 6.9e-05;  
Matches 176; Conservative 86; Mismatches 270; Indels 335; G  
QY 32 SP-GLATNSGDGETTTSAILLASVKELQPERLTREAEARQIVASQLERCXKLG  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 34 SFGVSTSSSDG-KAEKSQTAKKARVEASTPKVNKGGRSE-EISESEETNPAP  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 91 SMSMSASAEQFOWQDQCKDIIEFLTGLVDSCIRSLOESGILDPDYSTGC  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 92 TEOLPRPQSPLDSLDDG-----RSLNDGGSDPRIDPDQDN  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 148 PSLLSQSAQLNS-----KPEGSFO-----  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 133 PSIYSGSVENSDSSGLSGQPARYHPPLFPSPQPPDSTRPQEAFEPHPS  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 171 SYHSNQ---TLALGETTS-----QLPARGTQARATQGFSQ---GTVSRAGHLAI  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 193 GXHAPMEPTSNMFQPCAPPHPHQLYPGTGCVLSGPMPKGGGAASSVGCPNI  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 219 APPP-----PPPPEPPAPSGLSFHLPDAPPAAAAALYYSSSTLPA  
DB |||:||||::|::|::|::|::|::|::|::|::|::|::|::|:  
QY 253 HPPEPTTPISVSSSGASGAPTKPTPTPVGGG--NLSPAPP---PANPFHYTNLPP.
```

14:26:21 2004

us-09-501-171a-4.rsp

-----GGSP-----AAPOGSPKLOGGSAPEGATYAAAPRGSPX 301  
NNASAPPLGLCAQPLGHLPSHAMGQMGGLPGPEKGTLPADPSHSLPASSAP 367  
SRLAKSYSTSSPI---NIVVSSAGLSPIRVS-----PPTVOSTISSPIHQ 348  
MRFPYSSSSSSAAASSSSSSSSSASFFPASQALPSYHSPPTSLSVSNOPPKY 427  
TIGTYATLS-----PTKRLVHAS-----369  
SLFSQAVWSQGPFPFPYGRLLANSNAHPGFPPESTGAQSTAHPVPVSTHHHHQOQ 487  
YKHSQZLYATATLORGS---LAASRASYSQCHG---LGPELRALQ---SPE 417  
QOQOQOHHGNSGPPPGAPHPLEGG---SSHHAHPYAMGSLGLRPPYPGPA 542  
--HIDPIYEDRVYQKPPMRSLQ-----SQGD-----PLPPAHT- 450  
PPHSQVSYQAGPNGPPVSSSSSSSSSTSQGSYPCSHPSQGPQAGAPYFPFPVPTV 602  
-----CTYRTSTAPS--SPGVDVSVPLQ 470  
SATLSTVIATVASSPAGYKTASPPGPPPYGKRAPSGAYKTATPPGYKPG--SPPSF 660  
SQHQPON---AAAATFORASYAAGPASNYADPYRQLOYCESVESPYKSGPA---- 522  
IPPGYRTSPPAGEFTPKGSPVTVGP-----GPLPPAGPSGLPS 702  
PEGTARSP--SIDSIQKDPREFGWRDPEL-----PEVIOMLOHQFPVSQS 568  
PPAAPASGPPLSATQIKQEPAB-EYETPESFVPPPARSPSPPKVVDV-----PSHAS 756  
AYLOHLCFGDNKTKAEIRROGGIQLLVDLLDHRMTEVHRSAACGALRNLYVKANDDN 628  
RPNKHLDRGFN-----SCA--RSDLYFVPLEGS 785  
LKNCGGIPALVRLRKTDTLEIRE 655  
KKRAD-----LVEKVRREAEQRARE 808

April 22, 2004, 12:13:33



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

in search, using sw model

til 22, 2004, 12:10:15 ; Search time 57 Seconds  
(without alignments)  
6780.874 Million cell updates/sec

-09-501-171a-4

77

4FARKPPGAAPLAMPVDPQ.....PYSELNYETSHYPASPDSSWV 1225

SUM62

oop 10.0 , Gapext 0.5

.7041 seqs, 315518202 residues

s satisfying chosen parameters:

1017041

hth: 0

hth: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

PREMBL 25:\*

sp\_archaea:\*

sp\_bacteria:\*

sp\_fungi:\*

sp\_human:\*

sp\_invertebrate:\*

sp\_mammal:\*

sp\_mhc:\*

sp\_organelle:\*

sp\_phage:\*

sp\_plant:\*

sp\_rodent:\*

sp\_virus:\*

sp\_vertebrate:\*

sp\_unclassified:\*

sp\_rvirus:\*

sp\_bacteriopl:\*

sp\_archaeap:\*

the number of results predicted by chance to have a  
than or equal to the score of the result being printed,  
id by analysis of the total score distribution.

#### SUMMARIES

Seq	Length	DB	ID	Description
1.6	1265	13	Q801N7	Q801n7 xenopus lae
1.2	1149	4	Q86W91	Q86w91 homo sapien
1.4	1168	13	Q8AXM8	Q8axm8 xenopus lae
1.1	983	13	Q7T071	Q7t071 gallus gall
1.1	907	13	Q9DFB2	Q9dfb2 xenopus lae
1.9	742	13	Q9DE61	Q9de61 xenopus lae
1.3	956	11	Q7TS93	Q7ts93 mus musculus
1.2	956	11	Q924L4	Q924l4 mus musculus
1.1	956	11	Q8C3U7	Q8c3u7 mus musculus
1.1	962	11	Q924L5	Q924l5 mus musculus
1.8	938	11	Q80XQ4	Q80xq4 mus musculus
1.7	945	11	Q8CHF8	Q8chf8 mus musculus
1.2	295	13	Q90X45	Q90x45 brachydanio
1.9	892	11	Q8BRF2	Q8brf2 mus musculus
1.9	892	11	Q924L2	Q924l2 mus musculus
1.8	898	11	Q8BQ36	Q8bq36 mus musculus

17	1325	20.8	898	11	Q924L3	Q924l3 mus
18	1271	19.9	444	11	Q8BK47	Q8bk47 mus
19	1267	19.9	765	11	Q80VQ2	Q80vq2 mus
20	1256	19.7	781	5	Q9NHP1	Q9nhp1 dros
21	1249	19.6	562	5	Q8TOH3	Q8toh3 dros
22	1245.5	19.5	610	4	Q96FS1	Q96fs1 homc
23	1090	17.1	859	13	Q8AXM9	Q8axm9 xen
24	1075.5	16.9	785	13	Q8AXN0	Q8axn0 xen
25	969	15.2	1254	5	Q9UJ08	Q9uj08 caen
26	836	13.1	451	4	Q8NAH4	Q8nah4 homo
27	796.5	12.5	296	11	Q8BVH1	Q8bvh1 mus
28	788	12.4	795	11	Q9CQ73	Q9cq73 mus
29	719.5	11.3	593	5	Q8WPK5	Q8wpk5 oiko
30	524	8.2	134	13	Q910B4	Q910b4 sal
31	522.5	8.2	326	11	Q80MW5	Q80mw5 mus
32	522	8.2	134	13	Q910B3	Q910b3 sal
33	520	8.2	134	13	Q910A5	Q910a5 sal
34	520	8.2	134	13	Q910A9	Q910a9 onc
35	520	8.2	134	13	Q910A7	Q910a7 sal
36	520	8.2	134	13	Q90VZ7	Q90vz7 par
37	520	8.2	134	13	Q90VZ6	Q90vz6 sal
38	517	8.1	134	13	Q910B1	Q910b1 sal
39	517	8.1	134	13	Q90VZ8	Q90vz8 onc
40	516	8.1	134	13	Q910B2	Q910b2 sal
41	515	8.1	134	13	Q910B0	Q910b0 onc
42	512	8.0	134	13	Q910A6	Q910a6 sal
43	512	8.0	134	13	Q910A8	Q910a8 sal
44	504	7.9	295	6	Q28875	Q28875 dos
45	286	4.5	91	4	O95645	O95645 homo

#### ALIGNMENTS

#### RESULT 1

Q801N7	ID	Q801N7	PRELIMINARY;	PRT; 1265 AA.
AC	Q801N7;	AC	Q801N7;	
DT	01-JUN-2003	(TREMELrel. 24, Created)		
DT	01-JUN-2003	(TREMELrel. 24, Last sequence update)		
DT	01-OCT-2003	(TREMELrel. 25, Last annotation update)		
DE	Similar to plakophilin 4 (Fragment).			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RA	Klein S., Strausberg R.;			
RL	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC048023; AAH48023.1; -.			
DR	InterPro; IPR008938; ARM.			
DR	InterPro; IPR000225; Armadillo.			
DR	Pfam; PF00514; Armadillo_seg; 4.			
DR	SMART; SM00185; ARM; 6.			
DR	PROSITE; PS50176; ARM_REPEAT; 3.			
FT	NON_TER 1			
SQ	SEQUENCE 1265 AA; 139253 MW; 01F7CB67BBAAE2FB CRC64;			

Query Match 44.6%; Score 2842.5; DB 13; Length 1265;  
Best Local Similarity 48.4%; Pred. No. 1.1e-173;  
Matches 622; Conservative 192; Mismatches 319; Indels 153; G;

QY	11	PLGAMPVPPQPS---SASEKTSLSPLNTSGDSETTTSA-ILASVKEQLQFI
Db	62	PAGGVPVPGQTSLMEALQKITQAAPS-NT---GMEPTTATNIIASVKEQLQFI
QY	67	RELEAFRIQVASOLEKCLGSETGSMSSMSAEQFQWQSQDQKDIEELTTGLEI
Db	117	RELEVERQIVANQLEKRCUGAESPSIASVSSTKSPFWRK---GDVPAGGSPKPRI



```

>PEGTIARSPSIDSIQKDPREFQWRDPPELPEVIMQLQHOPSPVQSNAAAAYLOHLCHF 579
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ADGGTTRSPSIDSIQKDPREFAWRDPPELPEVIMQLQHOPSPVQANAAAAYLOHLCHF 557
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>KAEIRQGGIQLVLLDHRMTEVHRSACGALRNLYVKGANDNKIALKNCGGIP 639
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>KMEVCRLLGGIKHLVDLDRHVEVOKNACGALRNLYVFGKSTDENKIAMKNVGGIP 617
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>LRKTTDLRELVTVGLWNLSSCDALKMPIIQDALAVLTNAVIIIPHSWENSPQLQ 699
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>LRKSIDAEVRELTVGLWNLSSCDVAKMTIIRDALSTLTNTVIVHSGWNNSSFD 677
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>QLHSSQVLNATGCLRNVSAGAEARRRMRCDGLTDALLYVIOQALSGSEIDSK 759
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>KFQTSVLNRTTTCGLRNLSAGAEARKQMRSCGELVDLSLLYVIHTCVNTSDYDSK 737
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VCILNLSVRLAAETSOQGMGTDELGLLCEANGKDAESSGCGWKKKKKK-- 817
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VCTLRNLSYRLELVQARLLGLDELDDLLGKESPKOSEPS-CWGKKKKKKRT 796
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>WDVGGLPDCABPPKGIQMLHPSIVKPYLTLLSECSNPDTLEGAGALQNLAAQ 876
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>WDVGGPICGLSKSPKGVEMLWHPVVKPYLTLLAESSNPATLEGSGAGSIQNLASQ 856
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>WYIRAAVRKEKGLPILVELLRIDNDRVVCATARNMALDVRNKLIGKYAMRD 936
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>WYIRAAVRKEKGLPILVELLRIDNDRVVSATARNMALDVRNKLIGKYAMRD 916
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>PGNNSNTASKAMDDTVTAVCTTLEHVTIKMENAKALRDAGGIEKLVGLSKS 996
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>PGNGGPS-----VLSDETAAALCCALHEVTSKMNENAKALADSGGIEKLVNITKG 971
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ISPKVYKAAASQVLNWMQYRDLRSLYKDGWSQYHFVASSSTIERDRQRPYSSRT 1056
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ISLKVYKAAQVLTNLQYRDLRSYKDGWNQNHFIITPVSTLERDREKSH----- 1026
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>W--RVSPNNRSASAPSPREMISLKERKTDYECTGSGNATYHGAKGHTSRKDAMT 1114
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>TNQMSPIIQSGSKSP----- 1049
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ISTLYRNSVGAPEAD-----IKHNOVSAQVPQEPSRKDYETVQPFQNSTRYDES 1170
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>-----IVISSYSPARQNRRLQHQLYYS--QDSSNRKNFDAYRLYQSPHSTEDP 1099
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VHHRPPASEYTMHLGLKSTGNVDFYSAARPYSELNYETSHYPASPDSWV 1225
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VHF-PASTDYSYQYGLKSTNTVDFYSTKRP-----SYRAEQYPGSPDSWV 1149
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
<ELIMINARY; PRT; 1168 AA.
<EMBLrel. 23, Created)
<EMBLrel. 23, Last sequence update)
<EMBLrel. 25, Last annotation update)
: (African clawed frog).
: azoa; Chordata; Craniata; Vertebrata; Euteleostomi;
: achia; Anura; Mesobatrachia; Pipidea; Pipidae;
: enopus.
: 5;
: N.A.
: Vanlandschoot A., Staes K., van Roy F.;
: ated proteins in Xenopus laevis.";
: (-1999) to the EMBL/GenBank/DBSJ databases.
: ; AA013695.1; -
: 08938; ARM.
: 0025; Armadillo.
: ; Armadillo seg. 5-
: .76; ARM_REPEAT; 3.
: 8 AA; 128678 NW; 229DF17F55AA760D CRC64;

```

```

Query Match 43.4%; Score 2769.5; DB 13; Length 1168;
Best Local Similarity 48.6%; Pred. No. 4.9e-169;
Matches 609; Conservative 180; Mismatches 338; Indels 125; G

QY 15 MPVPDPSSASRKTSLSPGLNTSNGDSGTETTSAILASVKEQELQFERLTRELE
Db 1 MPVPEQTPLMEERALKITQISAPSTGMEPTTATN-ILASVKEQELQFERLTRELE
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 75 IVASOLERCKLGSSETSMSSMAEFOFQWQSQDGQKXIDEDLTGTGLELYDS----
Db 60 IVANQLERCKLGAESISGVSTEXSPWRS-----GDVPAGGSPKPLSDSHPN
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 RSLQE-SGILDPQDYSTGERPSLL--SOSALQLNSKPEGSFOYPASVHNOTLALG
Db 116 RSEPEPMNAVYSPQSSUHEREGSMGNSRSQMNYSYSDSGYQAGSFHNN--IG
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 185 PSQLPARGT-----QARATGQSPSQG-----TTSRAGHLAGPEPAPPPPP----
Db 173 RAQLSFTGSTGNHAVRNSRAEGQTFVQPPNTMAGSRAVRRVSVSPRSQSPPSYVI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 225 -PPREPPAPSLGSAFHLDPAPPAALAAALYSSSTLPAPPRGGSLAAPQGGSPTK
Db 233 SPBGSRLRTSVGSGYSPSTSDPRPLATNYSSTLLPS-QRAASPYTLHRSNGETV
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 284 GSAPEGATYAAPRGSSPKQSPKSLAKSYST-----SSPINIVVSSAGLSPIRVT
Db 292 GSI-----NSRQLSNPNTMTSSYQTLGSSARVGSPL---TSSDAHSRVGLS
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 337 VQSTISSPIHQLSSTIGYVATISPTKRLVHASEQYKHSQELIYATATLORPQSLA
Db 336 -QAPMGSSP-----KRAGMTAVPQHTGTL-NDIDQEQQHYDIYERMVPRDPSL-
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 397 ASYSSOHGHLGPELRALQSPERHIDPIYEDRVYQKPPMRSLSSQSGDPLPAHGT
Db 388 SSTASQHSQGLQELRSVAVPDLOITPIYEGRTYSPVYRNANHAADLROGSHTAI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 457 TAPSSPGVDSVPLQRTGSHQGPQNAATAAFQASVAAAGPASNADPYRQYQPCPV
Db 447 -----GSGSGNLQRTMSQR-----CTLAVQRNNYALNTAANYAEPVRSQYRIS-
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 517 SKSGPALPEGLTLARSPSIDSIQKDPREFQWRDPPELPEVIMQLQHOPSPVQSNAAA
Db 495 NRIPTHTPPDDGATSPSIDSIQKDPREFAWRDPPELPEVIMQLQHOPSPVQANAAA
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 577 LCPGDNKIKAETRRQGGIOLLVDLDRHMTVEHRSACGALRNLYVKGANDNKIAL
Db 555 LCPGDHKTLEVCRMGTGTHLVLDLONKYLEVQKNACGALRNLYVKGSTDENKIAV
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 637 GIPALVRLRKTTDLRELVTVGLWNLSSCDALKMPIIQDALAVLTNAVIIIPHSQ
Db 615 GVPALLRLKKTSQDQVRELVTGLWNLSSCDVAKMTIIRDALSTLTNTVIVPHSG
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 697 PLODDRKIQLHSSQVLNATGCLRNVSAGAEARRRMRCDGLTDALLYVIOQALG
Db 675 SFDDDHKMKFQSSVLRNTTTCMRNLSSSGEEARKQMYCEGLVDLSLLFVIQTCVN
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 757 DSKTVENCVCILRNLSYRLAAETSQGHMGTDELGLLCEANGKDAESSGCGWKKK
Db 735 DSKTVENCICITLRNLSYRLELVQVQVRLMGINEMDG-FCKESPKSDSDS-CWGKKI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 817 K---SQDQWDGVGPLPDCABPPKGIQMLHPSIVKPYLTLLSECSNPDTLEGAGA
Db 793 KKKPPQEDLWDGVGPIPLGSLQSPKGVEMLWHPVVKPYLTLLAESSNPATLEGSGAGSI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 874 AGSWKWSYVIRAAVTRKEKGLPILVELLRIDNDRVVCATARNMALDVRNKLIG
Db 853 SAGNWKFAAYIRAAVTRKEKGLPILVELLRIDNDRVVSATARNMALDVRNKLIG
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 934 MRDLVHRLPGGNNSNTASKAMDDTVTAVCTTLEHVTIKMENAKALRDAGGIEKI
Db 913 MRDLVNLVPGNGGPS-----ILSDDTVAALCCALHEVSNKNMENAKALADTGGIEKI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Qy	502	PRQLOQYCPSVBS-----PY--SKSGPALPPE-GTFLA-----RSPSGIDS
Db	324	PHKGRGYDDPIETEMVEERIPYLHGSAAPLAQPERGSMASIDRLGKRKSPGIDS
Qy	544	EFGWRDPPELPEVLOMLOHOFQSVQSNAAAYLOHLFCFQDNKIKABIRQGGTQLL
Db	384	---WRDPDPEVAMLSHPIDPVKSNAAYLOHLFCYENDKIKKQVRHLKGPIL
Qy	604	RMTVEHRSACGALRNLVYGKANDNDNKIALKNCGGIPALVRLLRKTTDLEIRELV
Db	441	PKPEVHRKACGALRNI SYGKN-ENKVAIKNCDGIPALIRLLRKTNDWEARELIT
Qy	664	LSSCALKMPIIQDALAVLTNAVII PHSGWENSPLODDRKIQLOHSSQVLRNATGC
Db	500	LSYSEPLQWVIINHGLQTLTNEVII PHSGWSENEPDSKPRDAEWTTVFKNTSGC
Qy	724	SAGEARRRMRECDGTLDAIVYQSLAGSSEIDSKTVENCVCILRNLNSYRLAAI
Db	560	SPGAEARRLRCDGLVDALLHALQSAVGKKDTNKSVENVCIMRNL SYHVHKI
Qy	784	HMGTHIELDGLLGEANG-----KDAESSGCWGKKKKKXSQDOWD---GVGPLI
Db	617	--GADYQOELDAGQTAGTGCGSKKKKDDAGCGKAKGKNGGLDLSRFTDLIL
Qy	835	PKXIGQMLHPSITKPYLLTLLSECSNPDTLEGAAGALQNLAAAGSKWKSIVYRAAVI
Db	675	AKGFELLQPDVVRVLYLSILTESQNFNTLEAAAGALQNL SAGNWTWSTYIATVVI
Qy	895	PILVELLRIDNRVVCVAVATARNALMDVNRKELI GKYAMRDLVHRLP DGGNNSNI
Db	735	PVLVELLOSDSKVVRVAVSIARLNSLRDRNRKDLIGSYANGELVRNLP---SRQC
Qy	955	MSDDTWTACCTLHEVITKWNENAKALRDAGGIEKLVGISKSGDKHSPKVVAKAI
Db	792	LEEDIVVAVLTIIHEIITDSENARSIIQTGIOKLVAISKS---SOSPRETHKAI
Qy	1015	SMWQYDRILRSLYKKOGWSQYHFVASSSTIERDR---QRPYSSSRTPSIPVRVE
Db	849	MTWSYKELRNALQKDGWNKSHFQSVSATPKSSKGTAGRSYDDSTLP LVD--KSC
Qy	1071	ASAPASPREMISL-----KERTDYECTGNSAIYHGKAGEHTSRK
Db	907	GS-----RDMIPMDELPGDGYSTTDHRRDKERKYYKTSNTGDASEKEPLKNDTNRK
Qy	1116	QNTGISTLYRNSYGAPAEIDIKHNVSAQPV 1145
Db	962	NRASYNLV-----DARDIK-----POPV 979

RESULT 5	Q9DFB2	PRELIMINARY;	PRT; 907 AA.
ID	Q9DFB2		
AC	Q9DFB2;		
DT	01-WAR-2001 (TRENBLrel. 16; Created)		
DT	01-WAR-2001 (TRENBLrel. 16; Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25; Last annotation update)		
DE	Catenin arvcf-2ABC protein.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI TaxID=8355;		
RP	[1] _SEQUENCE FROM N.A.		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=20459074; PubMed=10899158;		
RT	Paulson A.F., Mooney E., Fang X., Ji H., McCrear P.D.;		
RT	"arvcf: Xenopus Member of the p120 Catenin Subfamily Associatin		
RT	Cadherin Juxtaembrane Region."		
RT	J. Biol. Chem. 275:30124-30131 (2000).		
DR	EMBL; AF287051; AAG00555.1; -		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000225; ARMdillo.		

RESULT 6	Q9DE61	PRELIMINARY;	PRT;	742 AA.
ID	Q9DE61	PRELIMINARY;	PRT;	742 AA.
AC	Q9DE61			
DT	01-MAR-2001	(TrEMBLrel. 16, Created)		
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE	P120	(Fragment)		
OS	Xenopus laevis	(African clawed frog)		
OC	Eukaryota;	Metazoa;	Chordata;	Cranialata;
OC	Amphibia;	Batrachia;	Anura;	Mesobatrachia;
OC	Xenopodinae;	Xenopus		
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ciesiolka M., Vanlandschoot A., Staes K., van Roy F.;			
RT	"Armado-1-related proteins in Xenopus laevis."			
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF150746; AAC45945.1; -			
DR	InterPro; IPR008938; ARM.			
DR	InterPro; IPR000225; Armadillo.			
DR	Pfam; PF00514; Armadillo_seg. 4.			
DR	SMART; SM00185; ARM; 3.			
DR	PROSITE; PS01176; ARM_REPEAT; 3.			
FT	NON TER 1			
SQ	SEQUENCE 742 AA; 82927 MW; E51D3A3A5CB95FE7 CRC64;			
Query Match	22.9%;	Score 1459;	DB 13;	Length 742;
Best Local Similarity	45.3%;	Pred. No. 4.3e-85;		
Matches	342;	Conservative 105;	Mismatches 203;	Indels 102;
QY	389	GSIAAG--SRASYSSQHG-HLGPRLALQSP---EHHIDPIYEDRVYQKPPMRS--		
DB	11	GSLSRGINVRPRVAYTPGNRYRPEDGSYTLPIRRENYV-PIAQPV---PIGSNI		
QY	440	SQGDPLPPAHTG-----		
DB	66	SQPERFQPEYPLEDDNRSIGADDDGYELDDPDYSTWNRVLPG---MPL--TGPH		
QY	481	AAATFORASYAAGPASNVADPYRQLOCPSPVES---PYKSGPALP---PE-GTLA		
DB	117	-----KARAYDDPLE-----ADLIEERIPYLHGGYAAPLAQPERGSLA		
QY	531	---RSPSIDSIQKDPREFGRDPELPEVIOMLQHPSPVCSNAAAYLHLCFCGDNK		
DB	159	LGRKSPSIDSIKNDPR---WRDPDLPEVIAMLNHPIDPVKSNAAAYLHLCVENDK		
QY	588	IRPQGGIQLLVLDLDRMTVEHRSACGALRNLYVGKANDDNKIALKNCGGIPALVRI		
DB	216	VRYLKGIPTLVGLLDHPKPEVHRKACGALRNISYKDN-ENKVAIKNCGIPALIRI		
QY	648	TTLEIRELVTGVLWNLSSCDALKMPIIODALAVLTNAVIIIPHSGWENSPLODRK		
DB	275	TNDMEVRELTITGLNLSSEYEPKWIINHGLQTLTNEVIIIPHSGWENEPNEDSKP		
QY	708	SSOVLANATGCLRNVSAGFEARRRRECDGLTDALLYIQSALGSSESDSKTVENK		
DB	335	WSVFNKTSGLRNVSDDGAEARRRRECDGLVDALLHALQSAVSKKOTDNNKSVEN		
QY	768	LRNLSYRLAAETSSQGHMGTDELD---GLLCCGEANGKDAESSGCWKKKKKKKSDQ		
DB	395	MRNLSYVHVKEVPGADRFQENELNQGGI---GGQKKKKDDAGCFGGKKKAKKNGI		
QY	824	--GVGLPPDCAEPKGIQMLMHPISIVKPYLTLLSECSNPDTLEGAGALQNLAAGSV		
DB	453	NFDTLDLPKRSSESAKGFELLYQEVVRYLYSLTLESQNTYLEAAGALQNLASGN		
QY	882	VYTRAAVRKEGLPTLVELLRIDNDRVVCATAVATARNMALDVNRKELICKYAMRDLA		
DB	513	MYTRAVRKERGLPVLVELLQSDSKVVRVAIAIARNLSMDRRNKDLTGNVAMGOLV		
QY	942	PGNNNSNTASKAMSDDTVTAVCCTLHEVITKNMENAKALRDAGIEKLIVGISKSK		

14:26:21 2004

us-09-501-171a-4.rspt

[illegible]

[illegible]

3ILASVKEQBAERUTRALQERRHVALQLERAQPGQMSGGWVG--SGOPLPMWQ 66

3OKDIEDELTGLELVDSCIRSLQESGIILDPQDYSTGERPSSLQ----SALQLNSK 161

3LVLQESPGSQASLATMPEAEVLEVTI 96

3QYPASYHSNQTILAGETTPSQLPARGTCARATQSGSQGTTSRAGHL--AGEPAP 220

3PGTPTSHVSIIVTSBDGTT-----RRTETKVT--KTVKTVTTTRVQVPLGSDGPP 148

3PPPPPPPREPAPSLGSFAHLPDAP-----PAAAAAALYSSSTLPDPP 263

3GPP-----LGS---PADGPLDRHYLLRGGGGPAATLSRYHSS-----187

3PLAARQGSPTKLQGGGAPAGATVAAPR-----GSPKQS-----PSRLAKSYSTSS 315

3GFPDGPESRDIPSYGSLRGVLGVRPRTGLLPGPDGCTFLGR--REAPMGS 242

3UVSSAGLPIRVTSPTVQSTTSSPIHOLSSITIGYATLSPTKGLVHASEQYKSH 375

3GFPSPGSL-----PEHFQAEPYG-----LEDDTSRLAADSGGPD 279

3YATATLQRPGLAAGRSASYSSQHGLGELPAALOSPEHHIDPIYEDRVYQKPPMR 435

3YSTATRRRP-----EYGR--GLRAFAFEDTADAGELIEER-----317

3SGODPLPBAHTGTYRTSTAPSSPGVDSVPLQRTGSHGQFNQAAAAATFORASYAAGP 495

3P-----FPAATPLAQ-----PERUGLSLR-----340

3ADPVRQLQYCPVESPYSGKGPALPEGTGLARSPIDSTQKDPREFGWRDPELPEV 555

3VVRSSPSVDSRKEPR---WRDELPPEV 365

3QHPFSPVSGNAAAYLQHLCFGDNKIKAETIRROGGTQLLVDLLDHRNTEVHRSACGA 615

3RHTVDPVKANAAYLQHLCFENEGIKRRVQRGLGJPLLVALLDHPRAEYVRREACGA 425

3VYGKANDDNKIALKXCGGIPALVRLRKTTDLIEIRELVTGVLWNLSSCDALXMPHII 675

3SYGR--DTDNKAARDCGGVPALVRLRAARDNEVRELVTGTLWNLSYEPLKXVII 484

3AVLTNAVILPHSGWENSPLODDRKQLHSSQVLRNATGCLRNVSSAGEBARRMRE 735

3QTLTHEVIVPHSGWREPNEDSKPDAEWTTFKNTSGCLRNVSSDGAEBARLRE 544

3TDALLYVTCQALSGSEIDSKTVCNCILRNL--SYRLAETSQOQHMGTD-----LDGL 793

3VDALLHAQSAEGRKDTNKSVCNVCINRNLISYHVKEVPGADRYQEAEPGIPGS 604

3ANGKDAESGGCWGKKKKKXQDWD---GVGLPDPACBPPKGIQMLHPSIVKPY 850

3RRRKD--DASCFGGKAKGKKAEDNRNFDTLDPKRTAAAGFELLYQPEVVRLY 662

3SRECSNPDTLEGAGALONLAAGSKWWSVYIRAAVRKEKGLPILVELLRIDNDRVVC 910

3TESRNFNTLEAAGALONLSAGNWTATYIRATVRKERGLPVLVELLQSETDKVVR 722

3IALRNMALDVRNKEGLICKYAMRDLVHRLPGNNNSNNTASKMSDDTIVAYCCTLHEV 970

3IALRNLSLDQRNKDLIGSYAMTELVRNVKQAQAPAH--PSAHEEDTVVAVLNTIHEI 781

3NMENAKALRDAGIEKLVGISKGDKHSPPVKAASQVLSNMWQYDRDLRSLYKDG 1030

3SLDNARSLLQARGVPALVALVAS---SQSVREAKAAASHVLQTVMSYKELGALQRDG 838

3YHFVASSSTTERDRQRPYSSTRPSPISPV--RVSPNNRKSASAPSPRE-----M 1080

3SRP--QSNASTAKPGKGTSSSGFDDSTLPLVDKSLDGEKSNTRDVI PMDTLGPDGAT 897

KERKTDYEYCNSATYHGAKEHTSRKDM 1113

3RRRT-----LGSdst-----GD-TSEKELL 920

RESULT 10	Q924L5	PRELIMINARY;	PRT;	962 AA.
ID	Q924L5			
AC	Q924L5			
DT	01-DEC-2001	(TrEMBLrel. 19, Created)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE	ARVCF isoform A1.			
GN	ARVCF.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mu			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6;			
RA	Saint-Jore B., Puech A., Merscher S., Xu H., Kuchterlapati R.,			
RA	Skoultschi A.;			
RT	"Developmental expression analysis of Arvcf, a candidate gene fo			
RT	velo-cardio-facial syndrome."			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF286212; AAK4214.1; -.			
DR	MGI; MGI:109620; Arvcf.			
DR	GO; GO:0005737; C:cyttoplasm; IDA.			
DR	GO; GO:0005634; C:nucleus; IDA.			
DR	GO; GO:0005886; C:plasma membrane; IDA.			
DR	GO; GO:0005515; F:protein binding; IPI.			
DR	InterPro; IPR008938; ARM.			
DR	InterPro; IPR000225; Armadillo.			
DR	Pfam; PF00514; Armadillo_seg; 4.			
DR	SMART; SM00185; ARM; 5.			
DR	PROSITE; PS00176; ARM_REPEAT; 3.			
SQ	SEQUENCE 962 AA; 105022 MW; 17B618AB987E1C55 CRC64;			
Query Match	22.1%;	Score 1411;	DB 11;	Length 962;
Best Local Similarity	34.5%;	Pred. No. 7.7e-82;		
Matches 388;	Conservative 148;	Mismatches 325;	Indels 264;	
Qy	47	TTSAILASVKGEQLFELTLRELAERQIVASOLERCKL-GSETGSMSSMSAAEE		
Db	8	SAASILASVKGEQAEFFELTLAQEQRRHVALQLERAQDFGMSGGVMVG-SGQPL		
Qy	106	SDQGQKDELTTCLEILVDSICRSLQESGILPDQYDSTGERPSLLSQ-----SAL		
Db	67	Q-----LVLQEQSPGQASLATMPEAPEVL		
Qy	162	PGSFQYPASVHNSOTLALGETTSPQLPARGTQARATGQSFSGTTSRAGHL-A		
Db	97	VVEDCGTPTSHVISIVSDGTT-----RTETKVT-KIVKTVITRTVQVPLG		
Qy	221	-----PPPPPPPREPAPSLGSAFHLDPAP-----PAAAMAAALYYSGST		
Db	149	LLDGGPP-----LGS---FADGPLDRHYLLRGGGPAAATLSRTYHSS--		
Qy	264	RGGSPLAAPQGSPTKYLQGSAGEGYTAAPR-----GSSPKQS-----PSRLAKS		
Db	188	-GG---GFPDGPESRDIPSYGLSRG;GVRPPTGTLGPGFGDGCFTLPGR-REA		
Qy	316	PINIVSSAGLSPIRVTPPTVQGTSSPHQLSSITGTYATLSPTKRLVHASE		
Db	243	E-----SQPPSGRL-----PEHQAEPYG-----LEDDTSLAADD		
Qy	376	SQLYATATLQRPGLAGSRASVSYSHGLGPELRALQSPHHIDPIVEDRVYQ		
Db	280	LEPDYSTATRRPP-----EYGR-GLRARAFETADADGELIEER----		
Qy	436	SLSQSGDELPAAHTGTVTSTAPSSFGVDSVPLQRTGSHGQFQNAATAATFORAS		
Db	318	-----PP-----PFAATAPLAQ-----PERGSLGLDR--		
Qy	496	ASNYADPVRQLQYCPVSESPYSKSGPALPPEGLTARSPSIDSTQKDPREFGRWRD		



RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield
RA	Krzywinski M.I., Skalska U., Smalius D.E., Schnerch A., Schein J.
RA	Jones S.J., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length
RT	and mouse cDNA sequences";
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16901 (2002).
RL	[2]
RL	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6; TISSUE=Brain;
RC	Strausberg R.;
RL	Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL	[3]
RL	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6; TISSUE=Brain;
RC	Strausberg R.;
RL	Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC043108; AAH43108.1; -
DR	EMBL; BC054544; AAH54544.1; -
DR	InterPro; IPR008938; ARM.
DR	InterPro; IPR000225; Armadillo.
DR	Pfam; PF00514; Armadillo_seg; 4.
DR	SMART; SM00185; ARM; 7.
DR	PROSITE; PSS0176; ARM_REPEAT; 3.
DR	Hypothetical protein.
KW	SEQUENCE 938 AA; 104924 MW; 2F13DB7350355832 CRC64;
SQ	
	Query Match 21.8%; Score 1390; DB 11; Length 938;
	Best Local Similarity 33.5%; Pred. No. 1.7e-80;
	Matches 374; Conservative 161; Mismatches 336; Indels 246; G
QY	41 DGSSTETTSAILASVKEQELQFERITRELEARRQIVASQLERCKLG-SETGSMSSM
DB	2 DDSEVESTASILASVKEQAEQEKLTALREERERVSQAQLERVRVSPQDANSLMAN
QY	100 EQFQWQSDQKDI EDELITGL ELVDSCIRSLQESSGILDPOD-----YSTGER----
DB	62 RRHQNGRFVGDALERKQKTSDLKL-----NG---PDQNHLLYSTIPRMQE
QY	152 SOSALQLNSKPSGFOVPASYHSNOTLALGETTSPQLPARGTQAR--ATGQSPSQG
DB	109 VETYTE--EUEPGA-----MSVSVSEITDD-----GTTTRTETTIVKKVKVT
QY	210 AGHLAGEPEPPPPPPPPREFFAPSLGSAFHLDPADPAAAAAALYYSSSTLPAPPRG
DB	152 T-----VQVPV-----NG---PDGLPVDASAVSNNYIQTLL-----
QY	270 AAPGGSPTKLGGGSAPEGATVAPRGSSPKQSPSLAKSYSTSSPINIVSSAG
DB	179 ----GRDFRKNNGGPGP-----
QY	330 RVTSPPTVQSTIGSSPIHQLSSTIGTYATLSPTKXLHVASEQYKSHSOELYATATL
DB	193 -----YVQAGTATLPRNFHYPPDGYGEHYEDGY-----
QY	330 SLAAGSRASYSSQHGHLGP-----ELRALQSPHHIDPIYERDVYQKPMRSLSSQ
DB	224 G-----SDNYGSLSRVTRIERYRPSMEGYRAPSRQD--VYGPQPVVRVGGSS
QY	446 PPAHTGYRSTAPSPGVDV-----PLQTSQHGHPQNAATAATFORASYA
DB	273 HRFPPEPYGLDDQRMGNYDDLDYGMMSDYGTARETGP-----
QY	497 SNYADPTVQLQYCFSPVESPYSKGPALEP-----GTLARSPSIDSIQK
DB	312 ----SDPARRLR-----SYEDMIGEVPDPQYVWAPLAQHERGSLA--SLDSLRK
QY	544 BFGWRDPELPEVIQMLQHPSPVQSNAAYLOHLCFGDNKIKAETIRQGGIQLLVD
DB	360 PSNRQGPPELPEVIAMLGFRLLDAVKSNAAAYLOHLCYRNDKVKTDVRKLKGPILVWG
QY	604 RMTVEHRSACALRNLYVGKANDDNKIALKQCGGTPALVLLRKTKTTLRETLVWG



```

RELIMINARY; PRT; 295 AA.
TREMBlrel. 19, Created)
TREMBlrel. 19, Last sequence update)
TREMBlrel. 25, Last annotation update)
(Novel gene similar to human, rat and mouse CTNND2
a 2, cadherin-associated protein delta 2, neural
elated arm-repeat protein) (fragment).
erio (Zebrafish) (Danio rerio).
tazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
i; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
anio.
55;
N.A.
T-2001) to the EMBL/GenBank/DBJ databases.
7; CAC94831.1; -.
008938; ARM.
00225; Armadillo.
; Armadillo_seg; 3.
5; ARM; 3.
176; ARM_REPEAT; 2.
1
1
95
5 AA; 32352 MW; 995237FAA98E6A09 CRC64;
21.2%; Score 1353.5; DB 13; Length 295;
arity 86.7%; Pred. No. 6.3e-79;
onservative 16; Mismatches 19; Indels 5; Gaps 1;
GVDSVPLQRTGSHQPGQNAATAAATQFASAAAGPASNYADPYQIQYCPSPVESPYSK 518
GVDSVPLQRTASQN-----ATGTPRGYASGQSANTYDIYRTLPYCSPVESPYSK 55
LPPEGTIARSPSIDSIQKDPREFWRDPELPEVITOMLQHPFSPVQSNAAAYLQHLCL 578
LPPEGNLARSPSIDSIQKDPREFWRDPELPEVITOMLQHPFSPVQSNAAAYLQHLCL 115
KIKAEIRROGGIQLLVLLDHRMTEVHRSACGALRNLYGKANDNDKIALKNCGGI 638
KIKAEIRROGGIQLLVLLDHRMTEVHRSACGALRNLYGKANDNDKIALKNCGGI 175
RLLRKTTDLREIRLVTVGLNLSLSSCDALKMPIIQDALAVLTNAVIIIPHSGWENSPL 698
RLLRKTSDEVREIRLVTVGLNLSLSSCDALKMPIIQDALAVLTNAVIIIPHSTWDVSPH 235
KIQHSSQVLRNATGCLRNVSAGAEARRRMRECEGLTDALLYVIQSLGSSSEIDS 758
KIQMHTSQVLRNATGCLRNVSAGAEARRRMRECEGLTDALLYVIQSLGSSSEIDS 295
RELIMINARY; PRT; 892 AA.
TREMBlrel. 23, Created)
TREMBlrel. 23, Last sequence update)
TREMBlrel. 25, Last annotation update)
eat gene deleted in velo-cardio-facial syndrome.
(Mouse).
tazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
heria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
090;
N.A.
6J;
683; PubMed=12466851;
nsortium,
ome Exploration Research Group Phase I & II Team;
the mouse transcriptome based on functional annotation of

```

```

60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
EMBL; AK044982; BAC32169.1; -.
DR MGI:109620; Arvcf.
DR GO:0005737; C:cytoplasm; IDA.
DR GO:0005634; C:nucleus; IDA.
DR GO:0005886; C:plasma membrane; IDA.
DR GO:0005515; F:protein binding; IPI.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR00225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR PROSITE; PSS0176; ARM_REPEAT; 3.
SQ SEQUENCE 892 AA; 97164 MW; 937AA47D0604ED90 CRC64;
Query Match 20.9%; Score 1334; DB 11; Length 892;
Best Local Similarity 37.2%; Pred. No. 6.1e-77;
Matches 352; Conservative 147; Mismatches 300; Indels 148; C
QY 223 PPPRPPFAP-SLGSFHLDPAPPAALYSSSTLPAPPPGGSPLAAPQGGSI
DB 2 PAELRQSPGSGASLATMPEAP-----EVLEETVVEEDP-----GTF
QY 282 RGSAPGATYAAPRGSPKQSPRLAKSYSTSPINIVSSAGLSPIRVTSPTT
DB 45 SIVTSDGTT-----RRTEK--VTKVTVTTRTVRQVPLGPDGL-PLDGGPPL-
QY 342 SSSPI---HOLSSITGVATLSPTKRLVHASEQYSKHSOELYATATLOR-----
DB 96 ADGPLDRHYLLRGCGGPAATLSRYHSSGGGPPGPPESRDIPSYGSLSRGLGVRI
QY 388 ----PGS---LAAGRSY--SSQHGHLGPELRALQSPHHIDPIYEDRVYQKPP-
DB 156 LGPGGPGDCTLPGRRAFPMGSESG--PPSGRSL--PEH-----FOAEPY
QY 434 -MRSI--SQSQGDP-LPFAHTGTYTSTAPSPGVDSVPLQRTGSHQGPONAAATI
DB 202 DTRSLAADDEGGDLEP-----DYSTAT-----RRPEYG-RGLARAI
QY 491 YAAGPASNYADPYQIQYCPSPVESPYSKSGPALPPEGTIA-----RSPSIDSI
DB 244 DDAGELLEERPPF-----PAATAPL-----AQPERGSLGSLDRVVRSPSVDSTF
QY 544 EFGWRDELPVITOMLQHPFSPVQSNAAAYLQHLQCFGDNKIKAEIRROGGIQLVIT
DB 293 ---WRDELPVITOMLQHPFSPVQSNAAAYLQHLQCFENEGIKRVRQRLGRLPLV
QY 604 RMTVEHRSACGALRNLYGKANDNDKIALKNCGGI PALVRLRLKTTDLREIRLVTVG
DB 350 PRAEVRRACALRNLSYGR--DTDNKAAIRDCGGVPALVRLRLRAARDNEVRELVTG
QY 664 LSSCDALKMPIIQDALAVLTNAVIIIPHSGWENSPLQDDRKIQHSSQVLRNATGCI
DB 409 LSSYEPLKWIIDHGLQTLTHEVIVPHSGWEREPNEDSKPRDAEWITVFKNTSGCI
QY 724 SAGEARRRMRECEGLTDALLYVIQSLGSEIDSKTVENCVCILRNLSYLAAEI
DB 469 SDGAARRRLRECEGLVDALIHQSAVGRKDTDNKSVENCVCIMRNLSYHVHKEV
QY 784 HMGTDG--LDGLLGEANGKDAESSGCGWKKKKKSDQDWD---GVCPLPDCAEE
DB 529 RYQAEPIPGSTTSQRRRKD--DASCFGGKAKGKKDAEMDRNFDITLDPRTTEP
QY 839 QMLWHPISIVKPYLTLLSECSNPDTLEGAAGALQNLAAAGSMKWSYVIRAAVKEKGL
DB 587 ELLYQPEVRLYLSLLTESRNFNTLEAAGALQNLASAGNTWTATYIRATVRKERGL
QY 899 ELLRIDNDRVVCAVATARNMALDVNRKELIGKYMARDLVHRLPGGNNSNNTA--S
DB 647 ELLQSETDKVRAVAIALRNLSLDQRNKLTGSYAMTELVRN---RNAQAPHP
QY 957 DDTVTAVCCITHEVITKNMENAKALRDAGGLEKLVGISKSGDKHSPKVYKAASQV
DB 704 EDTVAVLNTIHEIVSLSLDNARSLLQARGVPALVALVAS---SQSVREAKAASHV

```

```

DLRSLYKKQGSQVHFVASSSTIERDRORPYSSSTPISPV--RVSPNNRASAP 1074
: : : : : : : : : : : : : : : : : : : : : : : : :
:ELRGALQRDQGWTKGRF-QSASTAKGPKGTSPSSGGFDDSTLPLVJDSLDEKGNTRD 819
: : : : : : : : : : : : : : : : : : : : : : : : :
EE-----MISLKERKTDYECTGSNATYHGAKGHESTRKODAM 1113
: : : : : : : : : : : : : : : : : : : : : : : : :
:MDTLGPDGYAVDRRRRT-----LGS DST-----GD-TSEKELL 856
: : : : : : : : : : : : : : : : : : : : : : : : :
PRELIMINARY; PRT; 892 AA.

(Tremblrel. 19, Created)
(Tremblrel. 19, Last sequence update)
(Tremblrel. 25, Last annotation update)
B2.

(Mouse).
Chordata; Craniata; Vertebrata; Euteleostomi;
Herpetia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
0090;

1 N.A.
6; Puech A., Merscher S., Xu H., Kucherlapati R.,
all expression analysis of Arvcf, a candidate gene for
facial syndrome."
to the EMBL/GenBank/DBJ databases.
5; AAK64217.1; -.
20; Arvcf.
37; C:cytoplasm; IDA.
44; C:nucleus; IDA.
46; C:plasma membrane; IDA.
5; F:protein binding; IPI.
3008938; ARM.
000225; Armadillo.
4; Armadillo_seg; 4.
35; ARM; 5.
1176; ARM REPEAT; 3.
22 AA; 97121 MW; 9B29550ACB598908 CRC64;

20.9%; Score 1331; DB 11; Length 892;
larity 37.2%; Pred.No.9.5e-77;
Conservative 147; Mismatches 300; Indels 148; Gaps 37;

PREPAP-SLGSFAFLPDAPPAALYSSSTLPAPRGSGSLAAPQGSPTKLQ 281
: : : : : : : : : : : : : : : : : : : : : : : : :
LREQSGPGCASLATWPEAP-----EVLEETVTEEDP-----GTPTSHV 44
: : : : : : : : : : : : : : : : : : : : : : : : :
SAPEGATYAAPRGSPKQPSRLAKSYSTSSPINIVSSAGLSIRVTSPTVOSTI 341
: : : : : : : : : : : : : : : : : : : : : : : : :
TSGDGT-----RTETK-VTKTVKVTTRVQVLPGLDGL-PLLDGGPPL--GSF 95
: : : : : : : : : : : : : : : : : : : : : : : : :
PI---HQLSSTIGTYATLSPTKELVHASEQYKHSQELYATATLQR----- 387
: : : : : : : : : : : : : : : : : : : : : : : : :
LDRHYLLRGGGGPAATLSRTHSSGGGFPDGPESRDIPYSGSLRGLGVRRPTGL 155
: : : : : : : : : : : : : : : : : : : : : : : : :
PGS-----LAAGSRASY--SSQHGHLGPELRAQLQPEHHIDPIYEDRVVQKPP----- 433
: : : : : : : : : : : : : : : : : : : : : : : : :
GPGDGCFTLPGREAFPMGSESG--PFSGRSL--PEH-----FQAEFYGLD 201
: : : : : : : : : : : : : : : : : : : : : : : : :
SL--SQSGQDP-LPPAHTGTVYRTSTAPSSPGVDSVPLQRTGSHQGPQNAATFORAS 490
: : : : : : : : : : : : : : : : : : : : : : : : :
SLAADDEGGPDLEP-----DYSTAT-----RRRPEYG-RGLRARAFEDTA 243
: : : : : : : : : : : : : : : : : : : : : : : : :
GPASNADVPYQIQCPSEVSPYSKSGPALPPEGTIA-----RSPSIDSIQDKPR 543
: : : : : : : : : : : : : : : : : : : : : : : : :
GELIEERPPP-----PAAATPL-----AQPERSGLSLDRVRRSSVDSTRKEPR 292
: : : : : : : : : : : : : : : : : : : : : : : : :
WRDPPELVETOMLOHOPSPVSQSNAAAAYLOHLCFDGNKIKAEIRROGGTOLLVLLDH 603

```

[illegible]

Search completed: April 22, 2004, 12:14:43  
Job time : 64 secs